THE FORMATION AND FUNCTION OF THE HISTONE LOCUS BODY IN HISTONE mRNA BIOGENESIS

Kaitlin Pauline Koreski

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology

Chapel Hill
2019

Approved by:

Bob Duronio

Bill Marzluff

Jeff Sekelsky

Brian Strahl

Zbigniew Dominski

ABSTRACT

Kaitlin Pauline Koreski: The formation and function of the Histone Locus Body in histone mRNA biogenesis
(Under the direction of Robert J. Duronio and William F. Marzluff)

The genome exerts spatial and temporal control of gene expression through the compartmentalization of nuclear space into specialized substructures known as nuclear bodies (NBs). NBs are defined by light microscopy as the concentration of factors involved in specific biological reactions. In concentrating reaction factors and substrates in a distinct microenvironment, NBs are postulated to promote the efficiency of their associated reaction. However, a complete appreciation of how NBs form is needed to understand how NBs contribute to their in vivo reactions. To understand the relationship between formation and function of NBs I used the *Drosophila melanogaster* Histone Locus Body (HLB) as a model. The HLB assembles at replication-dependent (RD) histone genes and contains factors involved in histone mRNA biogenesis (i.e. transcription and processing). The RD histone mRNAs are of only known eukaryotic RNAs that do not end in a polyadenylated tail but rather end in a conserved stem-loop. We defined critical sequences within the 300nt H3-H4 bidirectional promoter that are essential for HLB formation, histone expression, and recruitment of a zinc-finger DNA binding protein, CLAMP, that helps regulate the locus. I then used engineered histone locus transgenes and found that the cis sequences required for HLB formation are dependent on the presence of the endogenous histone gene locus. I demonstrated that the H2a-H2b promoter can nucleate HLB components but only in the absence of the endogenous histone genes. This work suggests a role

of multivalent interactions in the formation of the HLB. This work provides insights into how the HLB forms and how this formation is related to the HLBs role in coordinating the steps in histone mRNA biosynthesis

ACKNOWLEDGEMENTS

I am incredibly thankful for everyone that has encourage and supported me on the path to this degree. Thank you will never be enough. I would first like to thank my mentors, Bill Marzluff and Bob Duronio. By letting me be a member of both of their labs, I have had the amazing opportunity to work with two great men and become more knowledgeable and skilled in complementary areas of biology. They have let me explore my interests while keeping me from getting too far off the rails. More importantly, they have let me make mistakes, but they guided me in how I could fix them. They have helped me become a more confident scientist, a skill that is hard to learn. Bob, no one has ever challenged me the way you do, and I appreciate every moment of that. You have made me a stronger scientist. Bill, your knowledge of, everything, never ceases to amaze me and I have gained so much by just being around you. Thank you for always giving encouraging words when things got difficult with projects. I can honestly say that I do not know where I would be without your guidance, then and now.

Thank you to both the Marzluff and Duronio labs. My interaction with all of you over the years have helped make me into the scientist I am. You have supported me when graduate just got too hard.

I want to thank my friends both those that are still here and those that have graduated. Robin, you have made a better scientist, our talks on runs helped clarify things. You have kept me balanced throughout graduate school and have been there when things seemed too tough.

Thank you and remember, prove everyone wrong. I want to thank my family for supporting me along this long path. The journey is not over, and I thank you for never turning back.

PREFACE

Chapter 2: Histone locus regulation by the drosophila dosage compensation adaptor protein clamp. ***Leila E. Rieder*, Kaitlin P. Koreski*, et al.*** *(2017). Genes Dev. Jul 15;31(14):1494-1508* *******These authors contributed equally.*

Most of this work in this chapter is a previously published research article(Rieder et al. 2017). This work was done with Dr. Leila Rieder and Dr. Erica Larschan who made the initial discovery of CLAMP at the histone locus and its localization to the HLB. All the genomic/sequencing work was done by the Larschan lab in addition to the CLAMP mutants and embryo staining. I had previous shown (McKay et al. 2015), that HWT transgenes formed an HLB similar to the endogenous genes, with the exception of Mute (Fig. 2.1). This provided the basis of creating the GAGA mutant transgenes as we could assay their ability to form an HLB. These transgenes, and the deletion analysis in Supplemental 1 3, were created by a former postdoc associate in the Duronio Lab Dr. Kara Boltz. I assisted her in making this histone array. I created the HWT transgene used in the paper and performed the retargeting experiments, expression analysis, and immunostaining of salivary gland tissue. In this work we have determined that GA sequences in the H3-H4 promoter nucleate HLB assembly and CLAMP binds these sequences in early development to help regulate the histone locus and expression of the histone genes. It was reviewed and edited by Dr. Erica Larschan, Dr. Robert Duronio, and Dr. William Marzluff. Supplementary material can be found in the online version of this article.

Figure 2.1

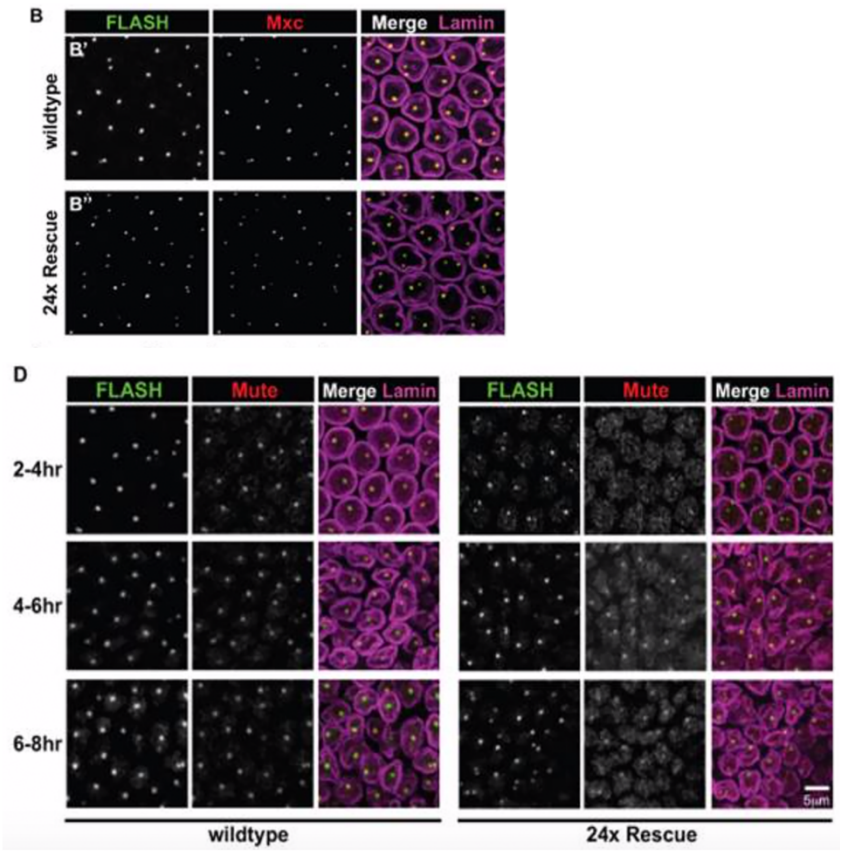


Figure 2.1. The transgenic histone gene locus assembles an HLB that accurately processes histone transcripts

. **(B)** Confocal images of blastoderm stage embryos stained for FLASH (green), Mxc (red), and Lamin (magenta) for the two indicated genotypes. **(D)** Confocal images of embryos at 2-4hrs, 4-6hrs, and 6-8hrs stained for FLASH (green), Mute (red) and Lamin (magenta) for wild type and 24× Rescue embryos. For **B** and **D**, the maximum projection of four 0.5-micron slices is shown.

McKay, D. J., S. Klusza, T. J. Penke, M. P. Meers, K. P. Curry, S. L. McDaniel, P. Y. Malek, S. W. Cooper, D. C. Tatomer, J. D. Lieb, B. D. Strahl, R. J. Duronio and A. G. Matera (2015). "Interrogating the function of metazoan histones using engineered gene clusters." Dev Cell **32**(3): 373-386

Chapter 3: Histone Locus Body Formation: Different ways to a common end.

This works represents a manuscript currently in preparation. My advisors, Dr. William Marzluff and Dr. Robert Duronio and I designed the experiments and analyzed the data. I performed all

the experiments with the exception of a few panels in Fig. 3.5C, of which my undergraduate, John Atwater, took a few images. My undergraduate, Lyndsey McLain, screened for transgenic fly lines and identified all 3 (HWT, PR, PR*) used in the study. I wrote the first draft of the manuscript which was then edited by Dr. Marzluff and later by Dr. Duronio.

Figure 3.1

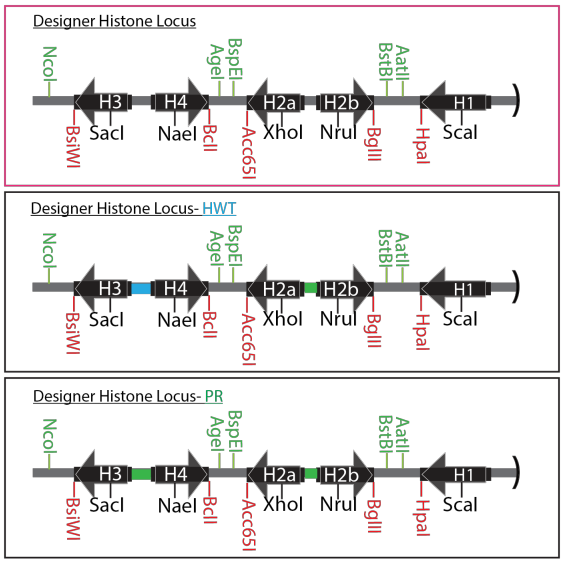


Figure 3.1 For this study I created a designer histone locus. This can be used to ask many questions regarding the regulation of histone gene expression, processing, and more. We designed it with unique restriction sites after each stop codon(red), after each histone downstream element (HDE) (green), and we made a silent mutation in the coding region of each gene by adding or destroying a restriction enzyme site. This enabled us to differentiate the expression from our transgene vs the endogenous genes. A version of this locus was used in the CLAMP retargeting studies in Chapter 2.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CB	Cajal Body
CLAMP	Chromatin linked adaptor for MSL proteins
CPSF	Cleave and polyadenylation specificity factor
EN	Endogenous genes
FISH	Fluorescent In Situ Hybridization
FLASH	FLICE-associate huge protein
GAF	GAGA factor
HCC	Histone Cleavage Complex
HCC	Histone Cleavage Complex
HDE	Histone Downstream Element
HLB	Histone Locus Body
HWT	Histone Wild Type
HWT*	Histone Wild Type that doesn't have the H3 gene silently marked
IDR	Intrinsically disordered region
LacI	Lac Repressor
LacO	Lac Operator
MSL	Male specific lethal
Mute	Muscle wasted
Mxc	Multi sex combs
NBs	Nuclear bodies
ncRNA	non coding RNA
NPAT	nuclear protein of the ataxia telangiectasia-mutated locus

nts.	Nucleotides
PR	Promoter Replacement
PR*	Promoter Replacement with one wild type repeat
RD	Replication-dependent
RRM	RNA Recognition Motif
RT-PCR	Reverse Transcriptase- Polymerase Chain Reaction
SLBP	Stem-loop Binding Protein
snRNP	small nuclear ribonuclear protein

CHAPTER 1: INTRODUCTION

OVERVIEW

Cells are faced with an important challenge: in the complex and crowded cellular environment cells must both spatially and temporally regulate thousands of simultaneous yet diverse molecular reactions to function properly. To accomplish this arduous task, cells display a high degree of compartmentalization which is thought to help regulate biochemical reactions. This compartmentalization is classically thought of as being achieved by numerous membrane-bound organelles such as the nucleus that sequesters our genetic information or the richly-shaped endoplasmic reticulum (ER). Membrane- bound organelles were beginning to be described in the 19th century. (Theory of Organelle Biogenesis: A Historical Perspective). Considering this, it is no surprise membrane-bound compartments provide our textbook understanding of intracellular organization. Despite this prevailing view of organization some of the first compartments to be described did not follow this paradigm (Shin and Brangwynne 2017). The nucleolus was formally described in 1898 (Pederson 2011) and the Cajal body was described in 1903 (Gall 2003) and we now know that these subcellular compartments are not surrounded by a membrane but rather freely exchange components with the surrounding environment.

In the nucleus spatiotemporal control over biochemical reactions is partially achieved through the formation of membraneless compartments known as nuclear bodies. Nuclear bodies (NBs) are microscopically defined by the concentration of factors involved in biochemical reactions, i.e transcription. By concentrating reaction factors, and excluding others, NBs are hypothesized to create specialized microenvironments that facilitate the efficiency of their

associated biological reactions. There is some evidence for this, but the function of NBs, and further why they form, remains largely unknown. To fully appreciate, and be able to test, the function of NBs, there is a need to have a thorough understanding of how they form. For my dissertation I sought to understand the relationship between NB formation and function.

To do this I have used a NB that forms at the replication-dependent histone genes as a model. The replication-dependent histone mRNAs are the only eukaryotic cellular mRNAs to be identified, even after multiple rounds of deep sequencing, that do not end in a poly (A+) tail, but rather end in a conserved stem-loop structure. The unique 3' end on the histone mRNA requires a specialized suite of factors to be properly processed and many of them are localized in the Histone Locus Body (HLB) (Liu et al. 2006). In this work, using the HLB, I addressed three related questions: (1) What contributions do NBs make to their in vivo reactions? (2) How do NBs specifically recognize their sites of function? (3) What cis acting elements contribute to NB formation?

In this chapter I will introduce our current understanding of nuclear body assembly and function. I will then discuss the Histone Locus Body, both the current ideas of formation and its role in histone processing. Finally, I will discuss using the HLB as a model to understand how NB formation is related to function.

Nuclear body assembly

In the nucleus, spatial and temporal organization of molecules and reactions is achieved by numerous membraneless compartments. These compartments are defined by their components and the reactions with which their components participate in, suggest a function. Many NBs have been identified to date including nucleoli, Cajal Bodies, promyelocytic leukemia

bodies (PML), histone locus bodies (HLB), and several others. NBs perform a diverse set of functions and are composed primarily of proteins and nucleic acids (RNA/DNA). NBs can be detected in the light microscope by using FISH probes to know sequences or antibodies against known components. It is not really understood how they concentrate components, control their composition, or influence their associated biochemical activities (Banani et al. 2017). In recent years there has been incredible strides made towards our understanding of these questions and with this have come the realization that misregulation of these processes can lead to devastating disease (Woulfe 2008).

It is attractive to look at the cell nucleus and observe many nonmembrane bound compartments and study their dynamics, composition, or function but the first challenge in understanding these compartments is defining what is required for the initial nucleation step that leads to formation. How do these molecules initially come together to give rise to what we see in the light microscope? Through years of biochemical and genetic analysis several major concepts of NB assembly have been proposed

Self-organization

An important organizing principle in cell biology is dynamic self-assembly, also referred to as molecular self-organization (Rajendra, Praveen, and Matera 2010). Put another way, this property says that a macromolecular complex determines its own structure based on the interactions between its components and further the interactions between its parts determine its function (Misteli 2001). When thinking about nuclear organization, as well as cytoplasmic, this property may seem familiar. Many components of NBs have the inherent ability to self-organize and as a result form visible macromolecular complexes but how this occurs has remined elusive.

One possibility is the stochastic, or random, assembly of components; the localization of one component does not depend on the other. Support for this assembly model was demonstrated by work from the Dundr Laboratory (Kaiser, Intine, and Dundr 2008) using a LacO/LacI tethering system to immobilize LacI tagged Cajal Body (CB) components to a genomic encoded 256 repeat LacO array. Subsequently, by staining for endogenous CBs components and screening for those that overlapped with the LacO array, they could determine the ability of CB components to nucleate the formation of a CB de novo. They found that the CBs could form by stochastic self-organization as assembly could be initiated by many CB components but only in the if CB components coilin and SMN (survival of motor neurons protein) were both present. This suggested that coilin and SMN are components that act cooperatively to facilitate CB formation. Conversely, it has been proposed that self-organization of components follows a hierarchal assembly pathway, with components associating in a defined sequence of steps (Dundr and Misteli 2010).

Seed and grow

Alternatively, assembly can occur as a combination of the two models. A single, or subset of components, may be required to initiate NB formation and continued assembly can occur via random localization of components. A study done by the Duronio lab (White et al. 2011) highlighted this potential hybrid model of NB assembly. Using a high throughout microscopy-based genome wide RNAi screen and genetic analysis, they identified Mxc and FLASH as HLB components required for localization of other HLB factors but not the vice-versa. This suggested stochastic self-organization does not tell the complete story of the HLB but rather implies that components within it show hierarchal self-organization. This work highlighted

a “seed and grow” concept of assembly where an initial nonrandom nucleation event occurs followed by stochastic self-organization assembles a NB. Many NBs form at the sites of transcription, suggesting that RNA may serve as an important seed in assembly. This has been shown for paraspeckle formation. Paraspeckles are NBs that contain long ncRNA species, Men ε/β (NEAT1), and various proteins. Paraspeckles are important players in the control of gene expression through nuclear retention of RNAs that have been subject to A-to I editing (Fox and Lamond 2010). Using a LacO/LacI tethering assay, much like the one mentioned previously, multiple paraspeckle proteins were able to recruit other paraspeckle proteins, however not very efficiently, but none were able to recruit the RNA associated with paraspeckles. Previous studies have shown that depletion of Men ε/β ncRNAs in cells disrupts paraspeckles indicating that these RNAs act as important structural component. Using a MS2 system to directly visualize transcription of Men ε/β ncRNAs and the recruitment of paraspeckle proteins it was found that transcription of Men ε/β was able to efficiently recruit all the paraspeckle proteins examined. This indicated that Men ε/β ncRNAs were the initial nucleating factor and provided a seed on to which to recruit other paraspeckle proteins stochastically (Mao et al. 2011).

Phase-separation

Recently a number of studies have shown that there are common features of membraneless compartments in both the nucleus and cytoplasm. For example, (1) there is a dynamic exchange of components with the surrounding nucleoplasm, or cytoplasm, (2) their structures are largely spherical, and (3) these structures can fuse and then relax into one spherical structure. These properties suggest that membraneless compartments behave like liquids (like oil drops in water) and form via a process known as liquid phase separation (Shin and Brangwynne

2017). Phase transitions are commonplace in nature when one state switches to another. Liquid phase separation can be seen in everyday life. For example, if two immiscible liquids are mixed, such as oil and water, this will always "demix" and separate into two distinct "compartments". Another example that is easily recognized is water. Water can exist in liquid, solid, or gas phases. In each of these phases the chemical composition of the water is the same, but the molecular organization is drastically different (Shin and Brangwynne 2017).

Phase separation usually occurs in a concentration dependent manner where there is a solubility limit, or threshold concentration, below which everything is mixed and once this limit is passed two phases exists; one in which is small and highly concentrated in a set of molecules and the other which is a low concentration dilute phase (Fig. 1.1) (Shin and Brangwynne 2017; Patel et al. 2015). This has been long observed in the process of X-ray crystallography. In supersaturated protein solutions, phase separation can occur resulting in two distinct phases with widely different concentrations of protein. Due to this phase separation, in the high protein concentration phase crystallization can occur much faster (Martin Muschol 1997).

Figure 1.1 Adapted from Taylor et. al 2016 (Taylor, Brown, and Cleveland 2016)

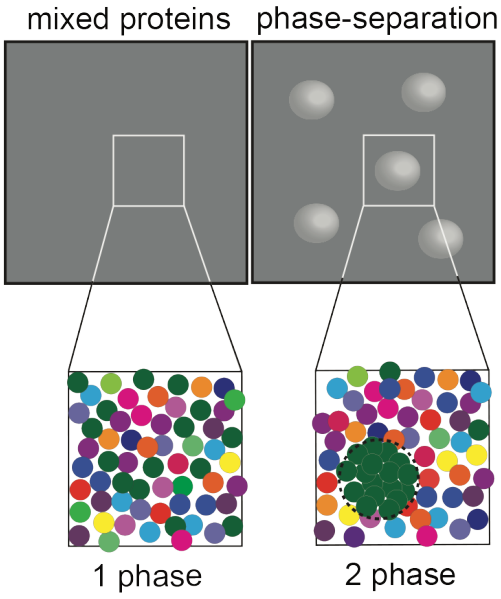


Figure 1.1 Schematic representing phase separation. Proteins exist in two phases- a dense phase and a dilute phase. On the left proteins are in a dilute phase. On the right, protein within the dilute phase transition to a more concentrated, dense state (green focus).

Even though phase-separation seems to have been recognized by structural biologists for some time it has only recently come to the stage as a common theme in organizing intracellular space. Membraneless compartments have been studied for a very long time but forces driving their formation have remained out of reach. The first step forward in this came with the observation that RNA and protein-rich P-granules in *Caenorhabditis elegans* displayed liquid like behavior. P-granules exist as a soluble phase and a condensed phase which is appeared to be spherical, when these granules attached to the nucleus, they became nonspherical displaying an appearance that resembled liquid droplets wetting a surface and moreover, P-granules occasionally fused. In addition, P-granules are dissolved and rapidly condense in the posterior of the embryos upon division. This suggested that P-granules may behave as a liquid and undergo phase separation (Brangwynne et al. 2009). Then, not long after, nucleoli were shown to have some of the same liquid- like properties (Brangwynne, Mitchison, and Hyman 2011). The number of membraneless compartments that display liquid-like behavior is growing but important questions still remain.

NBs are made up of many types of components (RNAs, DNA, proteins). In the bodies where they have been analyzed a small number of these components are required for the integrity of the body. These components are referred to as "scaffolds". An important property of scaffolds is the multivalent nature of the proteins, meaning that these proteins harbor multiple interaction motifs that drive intra- or inter- molecular interactions(Banani et al. 2017). These proteins include ones that harbor multiple modular interacting domains and proteins with stretches of intrinsically disordered regions (IDRs) and these provide multiple weakly adhesive elements which aid in the dynamic nature of NBs (Banani et al. 2016). Multivalent proteins are overrepresented in the composition of membraneless compartments and play an important role in

phase separation behavior (Feric et al. 2016; Mitrea and Kriwacki 2016; Li et al. 2012; Hyman and Simons 2012). Proteins are not the only thing that can provide multivalent interactions. Nucleic acids (RNA and DNA) which are often found in membraneless compartments can also contain multiple regions that bind to other nucleic acids and/or proteins. Together these scaffolds with multivalent interactions provide a mechanism to regulate the formation of these membraneless components.

Nuclear body function

Nuclear bodies form functionally distinct compartments within the 3D volume of the nucleus. The biological function of many NBs is known. For example, the nucleolus is the site of ribosome RNA biogenesis, CBs are involved in the assembly and modification of snRNPs (Mao, Zhang, and Spector 2011), paraspeckles are important for nuclear retention of RNAs that have been subject to A-to I editing (Bond and Fox 2009), and HLBs are involved in the transcription and processing of histone mRNAs (Marzluff and Koreski 2017) but how these bodies contribute to their in vivo reactions is not very well understood. By concentrating proteins and RNAs involved in specific biological reactions, NBs create distinct microenvironments that are postulated to increase the efficiency of these processes.

As mentioned, CBs play an important role in snRNP biogenesis but a detailed description about how they do this is still lacking. The pathway of snRNP biogenesis ends with the generation of a splicing competent tri-snRNP U4/U6•U5. In the first step of biogenesis U4 and U6 snRNAs are brought together by the combinatorial action of SART3 and Lsm2-8 proteins. SART targets U6 to the CBs allowing for annealing to occur in the CBs. After the U4/U6 duplex formation, proteins specific for U4/U6 are added creating the U4/U6 particle. This further

associates with the U5 snRNP and becomes the mature U4/U6•U5 tri-snRNP. Since the early discovery of coilin as a scaffolding protein required for formation of CB, the role CBs play in snRNP metabolism has been a matter of debate (Staněk and Fox 2017). Mathematical modeling and snRNP kinetic studies in suggested that snRNP assembly increased by a factor of 10 in CBs. This suggests that CB provide a cellular advantage in snRNP assembly. However, coilin depletion has been studied in *Arabidopsis thaliana, Drosophila melanogaster, Danio rerio* and *Mus musculus*. Loss of function mutations in coilin in plants and flies resulted in the dispersal of CBs yet, surprisingly, no major defects in viability or fertility were observed. In contrast depletion of coilin in zebrafish and mice had negative effects. In coilin -/- mice there was a dramatic effect on viability and fertility and in zebrafish embryos depletion of coilin was lethal within 24hrs and there was a reduction in snRNP levels and spliced mRNAs. This defect was rescued by the injection of assembled snRNPs into the embryo suggesting that the main function of coilin is to promote assembly of snRNPs (Machyna, Neugebauer, and Stanek 2015). These results highlight that disruption of a NB doesn't always have obvious impacts on the biochemical reactions with which it is associated. The functions of NBs are still a matter which is debated.

Figure 1.2

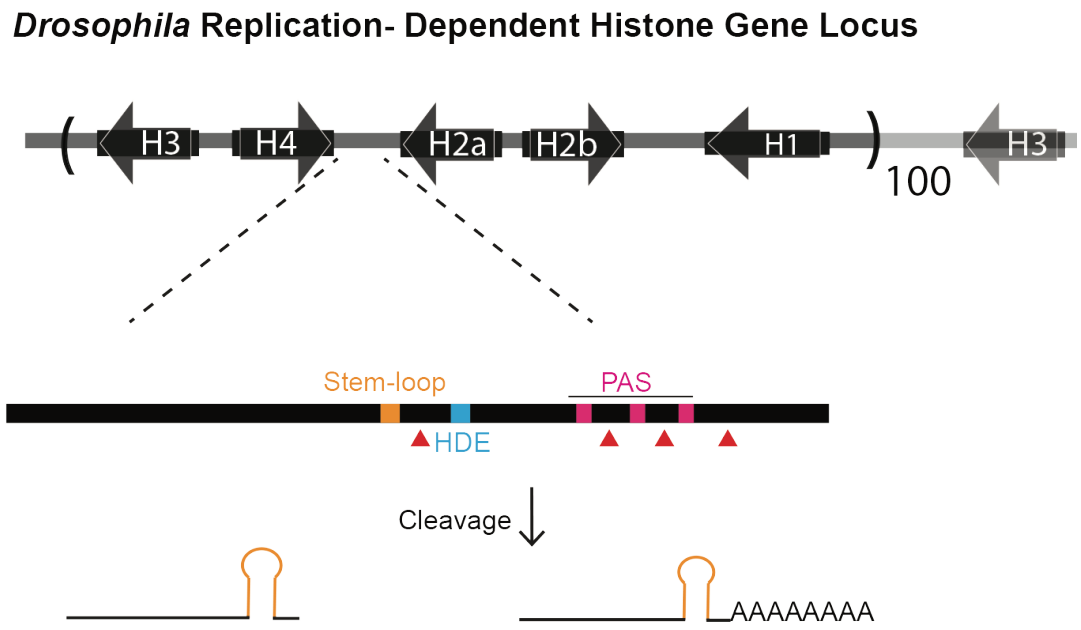


Figure 1.2. Drosophila Replication Dependent histone locus. These genes are present at a single locus as a tandemly arrayed 5kb repeat present in ~100 copies. Downstream of the histone processing signals (stem-loop and HDE), on all five histone genes there are cryptic polyadenylation signals (PAS). These are only used if the histone processing reaction doesn't occur efficiently and results in polyadenylated histone transcripts.

Histone locus body

Histone mRNAs and Components

Histone mRNAs are tightly regulated and present in high levels only in S-phase, to provide the histone proteins necessary for packaging the newly replicated DNA. The high demand for histone protein in S-phase, is met by the coordinated expression of multiple copies of the replication dependent histone genes. In metazoans all five replication-dependent histone genes have remained tightly clustered through evolution. This could reflect their presence in a specialized nuclear domain that creates a microenvironment for efficient histone mRNA

biogenesis. This is supported by the fact that in *C. elegans* the mechanism for 3'-end formation is different than the U7 snRNP dependent mechanism. This resulted in loss of the tight linkage of all the histone genes, and of the HLB

As mentioned above the replication-dependent histone genes are the only mRNAs that are not polyadenylated but instead end in a conserved stem-loop structure that is critical for their regulation (Fig. 1.2) (Pandey and Marzluff 1987). The stem-loop participates in all aspects of histone metabolism and is bound by the **S**tem **L**oop **B**inding **P**rotein (**SLBP**) which provides all the functions of the polyadenylated tail. The stem-loop and SLBP complex function in the processing (Lanzotti et al. 2002; Sullivan et al. 2001), transport(Sullivan et al. 2009), and translation (Cakmakci et al. 2008) of histone mRNA (Marzluff and Koreski 2017).

In addition to SLBP the formation of the unique 3' end of the histone message requires additional factors to be properly processed. Formation of the 3' end is mediated by two sites in the RNA; the stemloop and the histone downstream element (**HDE**). The stemloop is bound by SLBP and the HDE base pairs with U7 snRNP 3' of the cleavage site. The **U7 snRNP** is composed of U7 snRNA which is a small (<70 nt) RNA and a heptametric ring of Sm proteins that surrounds the U7 snRNA. 5 of the Sm proteins are those found in spliceosomal snRNPs: B, D3, E, F, and G and 2 proteins, Lsm10 and Lsm11, replace the spliceosomal proteins SmD1 and SmD2. Lsm11 is much larger than other Sm proteins, 360 aa in mammals and 256 aa in *Drosophila*, and the N-terminus of Lsm11 plays a critical role in histone pre-mRNA processing (Burch et al. 2011). This complex makes up the core U7 snRNP which is required for processing in *Drosophila* and mammals. The N-terminus of Lsm11 binds the N-terminus of another essential processing factor, **FLASH** (Flice Associated Huge Protein). FLASH was first identified as a pro-apoptotic protein and subsequently shown to be essential in histone pre-mRNA

processing (Yang et al. 2009). The N-terminus of FLASH interacts with the N-terminus of Lsm11 and together form a platform on to which recruit the **HCC (**Histone Cleavage Complex). The HCC is a complex of polyadenylation factors which includes CPSF73, the endonuclease that performs a single cleavage between the HDE and stemloop to generate the mature histone message and interesting enough. The identification of CPSF73 was a big surprise as this is the endonuclease that also cleave polyadenylated mRNA (Yang et al. 2013) Some of the factors discussed above, and others below, are found within the HLB.

However, initial studies identified this body as a specialized Cajal body. In mammalian cells it was observed that U7 snRNA was localized near the histone genes, and that it colocalized with the Cajal body marker protein, coilin, suggesting it was present in a subset of Cajal bodies (Frey and Matera 1995). The first protein discovered to localize specifically to histone genes was NPAT. NPAT was discovered as a cyclin E substrate that localized in a nuclear body near the histone genes in mammalian cells, and like U7 snRNA, is not present in other Cajal bodies so it was thought to be a specialized Cajal body (Zhao et al. 2000; Ma et al. 2000). It wasn't until Joe Gall's lab observed U7 snRNA localized in a nuclear domain that was separate from U85 and U2 scaRNAs, which are unique to the Cajal body. This distinct body was named the histone locus body (HLB) and was found to often lay close to the Cajal body. (Liu et al. 2006) Subsequently, the Marzluff and Duronio labs identified **Mxc** (multi sex combs) as the Drosophila ortholog of human NPAT(White et al. 2011).

There is an ever-expanding list of additional components of the HLB that have been and continue to be identified. Below I will discuss some of the defining members of the HLB. As mentioned above, **NPAT** and the *Drosophila* ortholog **Mxc** are scaffolding proteins in the HLB and are critical for HLB formation. They are both targets for cyclinE/cdk2 and phosphorylation

is required for activation of histone expression(Wei, Jin, and Harper 2003; White et al. 2007). **FLASH**, another defining member of the HLB, was originally identified as a factor required for activation of caspase 8 in apoptosis(Imai et al. 1999) and further shown to be an essential processing factor in mammals and *Drosophila(*Yang et al. 2009*)*. Using an EMS-mutagenesis screen to identify components required for myogenesis, **Mute** (muscle wasted) was identified as a protein that caused progressive muscle loss in *Drosophila* embryogenesis. Mute was observed to localize to the nucleus as a single prominent focus. This was shown to colocalize with known HLB components Lsm10 and FLASH, so it was identified as a component of the HLB. Further, it was shown that when Mute was knocked down there was increased levels of histone gene transcripts. Based on this observation it has been proposed that Mute serves as a negative regulator of histone gene transcription. (Bulchand et al. 2010) **YARP** (YY1 associated protein related protein) is a homologue of *Drosophila* Mute, binds specifically to NPAT and likely acts as a repressor as well. (Yang et al. 2014). The proteins outlined above are constitutive members of the HLB and once assembled, do not require ongoing transcription as they are present in G1, when histone genes are not active. There are many more proteins that have been identified in the HLB, some of which are only present during S-phase (e.g transcription elongation factor Spt6, or the scaffolding protein found in the HCC Symplekin).(Duronio and Marzluff 2017)

HLB assembly

As discussed above a critical component of nuclear bodies are scaffolding factors(Banani et al. 2016). These proteins and nucleic acids are necessary for the coherence of the body. In the HLB, genetic knockout studies or mutations(Terzo et al. 2015) in *Drosophila* have demonstrated that Mxc serves as a protein scaffold for assembly(White et al. 2011) and by this same criteria

the histone genes themselves are also a scaffold. Mxc's scaffolding role is exemplified by the binding of both FLASH and Mute/YARP to different regions of Mxc/NPAT's C-terminus (Yang et al. 2014). The domains at the C-terminus of Mxc that bind FLASH, are essential to localize it to the HLB. In addition, by knocking down Mxc in S2 cells there was a disruption in HLB formation (White et al. 2011)

An important feature of NBs is that they are composed of several different types of molecules and these must remain dynamic and in constant flux with the surrounding nucleoplasm. This can come from multivalent weak interactions between molecules which are necessary for formation/maintain the body. (Hyman, Weber, and Julicher 2014) Oligomerization of scaffolding proteins are commonly seen in NBs. This has been observed for Paraspeckles. Some paraspeckle proteins form homo-or hetero oligomers and associate, via their RRM, to RNA and when these interacting motifs are deleted there is a loss of paraspeckles (Mao et al. 2011). Similarly, this has been shown for Mxc. In the N-terminus of Mxc, two domains mediate self-interaction and have multivalent binding ability that could help oligomerize Mxc. This is important for Mxc function as changing 3 amino acids render Mxc unable to support HLB formation.

The studies on the mechanisms for HLB assembly have been done in both mammals and Drosophila, using different approaches and this resulted in two non-mutually exclusive pathways. Dundr and colleagues used a LacO/LacI system to study the ability of functionally related RNAs to form major NBs (e.g. HLBs, paraspeckles, nuclear speckles, and nuclear stress bodies) in mammalian cells. In this system a specific RNA was tagged with MS2 stemloops and immobilized on 256 genomically encoded LacO repeats through LacI-MS2 coat protein's interaction with the MS2 stem loops and subsequently LacO repeats. To assess if RNA could

nucleate HLBs, H2b RNA tagged with MS2 stem loops was tethered to the LacO array. NPAT and FLASH accumulated on the tethered histone RNA and this was dependent on the histone stem-loop and HDE as deletion or mutation, respectively, abolished de novo HLB formation. Additionally, tethering components involved in the expression (NPAT) or processing of histone mRNA (FLASH, Lsm10 and 11, SLBP, CPSF73, CPSF100, and CPSF30) resulted in formation of an HLB. This suggest that histone RNA or multiple factors involved in its expression or processing can nucleate an HLB. (Shevtsov and Dundr 2011) This supports a stochastic model of assembly wherein the order of assembly is of little matter.

Contrastingly, our laboratory has proposed a hybrid model of assembly (outlined above) where Mxc/NPAT and FLASH provided the protein seed for assembly. Evidence for this comes from: (1) when Mxc or FLASH was knocked down via RNAi HLB assembly was dramatically affected, suggesting a scaffolding role, (2) using tightly timed embryo collections Mxc and FLASH foci formed a cycle before histone transcription begins and the stochastic recruitment of additional HLB components is visible only one cycle later, and finally (3) in mitosis the HLB dissembles but small amounts of Mxc and FLASH remained chromosome bound supporting a "bookmarking" role. These data suggest a model in which both hierarchal assembly followed by stochastic self-organization builds the HLB (White et al. 2011), a "seed and grow" model. Considering this and the disparity in methodology, the finding of Duronio and Dundr do not need to be at odds but rather are complementary.

Continuing studies have identified how additional components in the HLB come together and these have shown that the domains for localization to the HLB are not the same as the domains for function. For example, the N-terminus of FLASH binds to the N-terminus of Lsm11 and forms a platform that is required for processing and this will recruit the HCC that is

required for proper processing (Burch et al. 2011; Tatomer et al. 2016a). Yet, the C-terminus of FLASH is required for its localization to the HLB. Further, U7 snRNP localization to the HLB requires the C-terminus of FLASH likely together with Mxc but does not require the region of Lsm11 required for processing.

We are beginning to understand total of interactions between the proteins in the HLB and how they come together and aid in assembly but the question regarding the initial nucleation or "seeding" event is still one that is not fully understood. The HLB invariably associates with the replication-dependent histone genes, in addition, as mentioned above, HLB components remain associated with the chromosome in mitosis, and once assembled, the HLB is present in G1 cells which do not have active histone transcription(White et al. 2007). This suggests that the histone genes themselves serve as the seed to initiate HLB assembly, and they also act as a scaffold for the HLB. To test this Duronio and colleagues have used a single histone repeat and various mutant forms to determine what, if any, part of the locus was capable of nucleating HLB components. A full-length histone repeat was capable of nucleating HLB and drive transcription of the locus. When just the H3/H4 or H2a/H2b gene pair were used, only the H3/H4 gene pair was capable of nucleating HLB components capable of driving histone transcription. This suggested that there was something different about the H3/H4 gene pair so to probe further into the requirements for HLB formation a promoter swap experiment was used. In this experiment, the H3/H4 coding regions were driven by the H2a/H2b promoter and the H2a/H2b coding regions were driven by the H3/H4 promoter. In this, the ability to nucleate HLB components and drive transcription followed the H3/H3 promoter. To investigate this more and to determine if any sequence in the histone pre-mRNA contributed to HLB assembly, just the H3/H4 promoter with no other histone sequence was tested. This nucleated HLB components and was capable of

drive transcription into the vector sequence. As I described above, the formation of Mxc and FLASH foci appeared prior to histone transcription up which further recruitment of components occurred(White et al. 2011). Considering this, to investigate if transcription from the H3/H4 promoter was required for full HLB assembly mutations in the TATA boxes in the H3/H4 promoter, preventing transcription, was assayed for HLB formation. This resulted in decreased recruitment of HLB components and no detectable transcription from not only the H3/H4 gene pair but also from the H2a/H2b gene pair. These data indicated that transcription from the H3/H4 promoter is necessary for not only full HLB formation but transcription from the H2a/H2b gene pair. This study identified the ~300nt H3/H4 promoter as a potential "seed" for HLB assembly and transcription from this is required for full HLB recruitment and expression for all replication-dependent histone genes in the locus(Salzler et al. 2013).

Dissertation goals

In my thesis project I have used the replication-dependent histone genes and associated Histone Locus Body as a model for nuclear body formation and function. It remains an open question as to whether condensates (NB and cytoplasmic compartments) provide a function, if any, to the cell. To fully understand this, rather than focusing solely on the output of the body there needs to be an appreciation of how and what is necessary for them form. These are important questions to understand as many cellular compartments have been implicated in neurodegenerative disease. Understanding the more about mechanism of condensate formation and how this formation is linked to function will provide a better understanding of what goes wrong when compartmentalization is affected in disease.

In the 2nd chapter I discuss work I did, together with Dr. Leila Rieder at Brown University, in defining the GAGA repeats found within the H3-H4 bidirectional promoter as critical sequences in HLB formation and expression. We identified CLAMP as the DNA binding protein that binds to the GA repeats within the promoter. Using the histone array technology developed in the McKay, Duronio and Matera labs(McKay et al. 2015), we found that when the GAGA sequences were mutated HLB formation and histone transcription were abolished. When CLAMP was tethered to an ectopic locus it could recruit additional HLB components. These results provided insight into how the HLB recognizes the histone locus

In the 3rd chapter I discuss my current work on the cis acting sequences that contribute to HLB formation and function. Using a number of synthetic histone gene arrays created to assess the contribution of the H3-H4 promoter to HLB formation and development, I found that the requirement of the H3-H4 promoter for HLB formation changes depending on the presence of the endogenous histone gene repeat. If the endogenous histone genes were deleted the H2a-H2b promoter nucleated HLB components. My results suggest a model of assembly where the H3-H4 promoter and the GAGA repeats are a higher affinity binding site for critical HLB components and sequester components away from the H2a-H2b promoter. This is consistent with HLB assembly being driven by multivalent protein-protein or protein-nucleic acid interactions.

CHAPTER 2: HISTONE LOCUS REGULATION BY THE DROSOPHILA DOSAGE COMPENSATION ADAPTOR PROTEIN CLAMP

INTRODUCTION

Within the complex environment of the nucleus, coordinated gene expression is facilitated by membraneless structures known as nuclear bodies (NBs). NBs are critical for the precise spatial and temporal regulation and processing of RNAs and include nucleoli, Cajal bodies, and histone locus bodies (HLBs) (Mao, Zhang, and Spector 2011). NBs share properties and assembly mechanisms with larger nuclear domains that regulate coordinated gene expression, such as the dosage-compensated X chromosome (in mammals, the Barr body). NBs improve the efficiency and coordination of nuclear processes, such as transcription and RNA processing, by concentrating factors to promote interactions that would otherwise be stochastic(Matera et al. 2009; Mao, Zhang, and Spector 2011; Tatomer et al. 2016a) . Despite their importance, our understanding of how specific NBs are formed early during development remains incomplete.

The HLB is a highly conserved NB that assembles at the replication-dependent histone genes(Liu et al. 2006), which are present in multiple clustered copies in most metazoans(Duronio and Marzluff 2017). Humans have two histone gene clusters, a major cluster on chromosome 6 and a minor cluster on chromosome 1(Albig and Doenecke 1997; Marzluff et al. 2002), while most *Drosophila* species have a single replication-dependent histone gene locus. In *Drosophila melanogaster*, the histone locus resides on chromosome 2L and consists of a tandem array of ~100 copies of a 5-kb cluster containing each of the histone genes[i]

[i]**Leila E. Rieder*, Kaitlin P. Koreski*, et al.** (2017). Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. Genes Dev. Jul 15;31(14):1494-1508 *These authors contributed equally

(Lifton et al. 1978; McKay et al. 2015). Coordinated expression of histone genes is necessary to maintain nucleosome subunit stoichiometry, and this requirement is reflected in the arrangement of the *Drosophila* histone genes that encode nucleosomal core proteins: In each 5-kb gene cluster, H2A and H2B share a bidirectional promoter, as do H3 and H4. This same arrangement is present in other species, such as budding yeast(Smith and Murray 1983; Eriksson et al. 2012). Histone production is also tightly coordinated across the cell cycle, leading to a burst of histone mRNA production at the beginning of each S phase(Marzluff, Wagner, and Duronio 2008). Many factors involved in the cell cycle-regulated transcription and processing of histone transcripts are concentrated in the HLB (Duronio and Marzluff 2017).

A common theme for NB assembly is that a "scaffolding" protein serves as a platform to recruit other NB components. In *Drosophila*, HLB scaffolding is mediated by the multi-sex combs (Mxc) protein, the ortholog of mammalian NPAT (nuclear protein of the ataxia telangiectasia-mutated locus), a Cyclin E/Cdk2 substrate that is essential for both HLB assembly and histone gene expression(Ma et al. 2000; Zhao et al. 2000; Ye et al. 2003; White et al. 2007; Terzo et al. 2015). Early during Drosophila development, before the initiation of zygotic histone gene expression, Mxc assembles into a "proto- HLB" along with FLASH (FLICE-associated huge protein)(White et al. 2011; Salzler et al. 2013) , a protein necessary for endonucleolytic cleavage to form mature histone mRNA(Yang et al. 2009; Burch et al. 2011; Tatomer et al. 2016a). Once Mxc and FLASH assemble into a proto-HLB, other factors involved in histone mRNA biosynthesis are recruited to the HLB (White et al. 2011; Salzler et al. 2013) , including the mRNA processing factor U7 snRNP (Strub and Birnstiel 1986; Mowry and Steitz 1987) and Mute (muscle wasted), a putative transcriptional repressor and homolog of the mammalian YY1-

associated protein (Bulchand et al. 2010; Yang et al. 2014). These data suggest that ordered recruitment of factors contributes to HLB assembly.

How the process of scaffolding the HLB is initiated and functionally linked to regulation of the histone locus chromatin and histone gene expression is not understood. Nucleation of Mxc/FLASH proto-HLBs does not require expression of histone mRNA(Salzler et al. 2013). Thus, one possibility is that a factor expressed during early development binds DNA at or near the histone genes and initiates HLB assembly and histone gene activation, perhaps by interacting with scaffolding factors such as Mxc/NPAT. Using engineered histone transgenes, Salzler et al. (2013) determined previously that the ∼300-base- pair (bp) bidirectional promoter between the Drosophila H3 and H4 genes (H3–H4p) is necessary and sufficient to recruit HLB factors, including Mxc, FLASH, U7 snRNP, and Mute. Although transcription from the H3–H4p is necessary for full recruitment of HLB factors, some Mxc and FLASH is recruited even in the absence of an active H3–H4p (Salzler et al. 2013). In addition, once fully formed, HLBs do not require ongoing transcription for maintenance, as they are present in G1 arrested cells that do not express histone genes(Liu et al. 2006; White et al. 2007). Thus, some HLB component likely recognizes a cis element in the DNA at the histone locus. The scaffolding protein Mxc contains one AT-hook domain, but there is no evidence that Mxc or NPAT directly binds DNA (Miele et al. 2005; Terzo et al. 2015; Wei, Jin, and Harper 2003).

The H3–H4p is highly conserved among 12 Drosophila species and contains two GA repeat cis elements(Salzler et al. 2013). GA-rich cis elements have been implicated in a variety of nuclear processes in *Drosophila*, including RNA polymerase II pausing (Tsai et al. 2016), zygotic genome activation(Chen et al. 2013), three-dimensional genome organization(Quinn et al. 2014), and DNA loop formation(Eagen, Aiden, and Kornberg 2017). Two known *Drosophila*

zinc finger transcription factors directly interact with GA repeats. The first, the well-studied GAGA factor (GAF; trithorax-like [trl]), opens chromatin and modulates transcriptional pausing at many genes(Guertin et al. 2012; Fuda et al. 2015). The second, chromatin-linked adaptor for male-specific lethal (MSL) proteins (CLAMP), is a zinc finger DNA-binding protein that is required for male X-chromosome dosage compensation (Larschan et al. 2012). CLAMP binds throughout the genome but is enriched at evolutionarily conserved long GA repeats on the X chromosome (Kuzu et al. 2016), where it recruits the MSL complex (Larschan et al. 2012; Soruco et al. 2013). The MSL complex generates a chromosomal domain of coordinated gene activation that increases transcript levels of male X-linked genes twofold, equalizing expression between XY males and XX females(Belote and Lucchesi 1980; Hamada et al. 2005). While not historically considered a NB, the male *Drosophila* X chromosome represents a distinct domain of coordinated gene activation similar to the histone locus.

Using genetic, genomic, and biochemical approaches, we show that the conserved GA repeats within the H3-H4p direct HLB formation. CLAMP, but not GAF, binds to these repeats early during development, before zygotic genome activation and prior to formation of the mature HLB. CLAMP is critical for histone gene expression and opening of chromatin at the histone locus. Furthermore, tethering CLAMP to an ectopic histone locus is sufficient to recruit HLB factors. Therefore, the presence of CLAMP and the absence of GAF at GA repeats at the HLB and the male X chromosome (Soruco et al. 2013) are common properties shared by two different domains of coordinated gene activation.

MATERIALS AND METHODS

Drosophila strains

We used the MTD (Bloomington, #31777) and a stock expressing a shRNA against clamp (Bloomington, #57008) made by the Transgenic RNAi Project (TRiP). For the H3–H4p deletion experiments we inserted promoter sequences into the pMulti-BAC vector containing a single histone repeat unit (McKay et al.2015) and inserted these transgenes into site 86Fb on chromosome3 using ϕC31-mediated integration (Bestgene) (Groth et al. 2004). The full sequences of engineered H3–H4p deletion sequencesare in the Supplemental Material. For the LacO array experiments, we synthesized H3–H4ps (Genescript) and used restriction digest cloning to insert the promoter containing LacO sequences in place of the wild-type promoter in the single histone repeat unit. We built an array of 12 histone repeat units in pMulti-BAC for each transgenic promoter and integrated each into site VK33 on chromosome 3 using ϕC31-mediated integration (Model Systems Injections). The full sequences of engineered H3–H4p with LacO sequences are in the Supplemental Material. We inserted CLAMPQ-LacI and LacI into the pUbi-GFP (gift from Mark Peifer), in which we swapped LacI for GFP using the LacI-HP1a vector (gift from Lori Wallrath). We amplified the CLAMP polyglutamine domain for Gibson assembly (New England Biolabs) using the primers F (5′-TAGGTCCTGTTCATTGAATGGAAGACCTTACCAAAAAC-3′) and R (5′-GTTACTGGTTTCACCATAGCCACAATTTGCTGAAG-3′). We drove transcription of both CLAMPQ-LacI and LacI genes using the ubiquitin promoter and integrated these transgenes into site VK20 on chromosome 3 using ϕC31-mediated integration (Genetivision). To make GFP-CLAMP, we cloned clamp cDNA into a vector containing the ubiquitin promoter (pUbi-GFP; gift from Mark Peifer) and integrated the transgene into site VK33 on chromosome 3.

Promoter alignment

We obtained promoter sequences from *D. simulans* (DNA Data Bank of Japan [DDBJ] accession no. AB055959) (Tsunemoto and Matsuo 2001), Drosophila erecta (DDBJ accession no. AB073634) (Kakita et al. 2003), *Drosophilia pseudoobscura* (DDBJ accession no. AB249651) (Nakashima et al. 2016), and *D. virilis*. We aligned sequences using T-Coffee (Notredame et al. 2000) and formatted the alignment using BoxShade.

FISH and immunofluorescence

We used primary antibodies at the following concentrations: rabbit anti-CLAMP (1:1000; Novus/SDIX) (Larschan et al. 2012), rabbit anti-CLAMP∗ (1:1000; custom antibody generated by our laboratory through a contract to Abcam; both anti-CLAMP antibodies were raised against the same N-terminal amino acids, CLAMP#22–121), guinea pig anti-Mxc (1:2000) (White et al.2011), guinea pig anti-Mute (1:5000) (Bulchand et al. 2010), rabbit anti-C terminus FLASH (1:2000) (Yang et al. 2009), rabbit anti-Lsm10 (1:1000), mouse anti-MPM-2 (1:100; Millipore), rabbit anti-GAF (1:1000; gift from Giacomo Cavalli), mouse anti-LacI (1:1000; Millipore), and chicken anti-GFP (1:400; Life Technologies). We used Alexa fluor secondary antibodies (Thermo Fisher Scientific) at a concentration of 1:1000. We detected in situ probes using 15 µg/mL streptavidin-DyLight-488 (Vector Laboratories). To make the FISH probe, we made a PCR product that spanned all five histone genes using a wild-type histone repeat in pMulti-BAC as the template (primers F [AAAGGAGGTTGGTAGGCAGC] and R [ACGCTAGCGCTTTATCTGCA]) (McKay et al. 2015).We made biotinylated FISH probes by nick translation using the purified PCR product: 1 µg of purified PCR product was incubated for 2 h at 15°C in a total of 50 µL containing 1×DNAPolI buffer (Fisher Optizyme); 0.05mMeach dCTP, dATP, and dGTP; 0.05 mM biotin-11-dUTP (Thermo Scientific); 10 mM 2-

mercaptoethanol;0.004 U of DNaseI (Fisher Optizyme); and 10 U of DNAPol I (Fisher Optizyme). The reaction was purified on a PCR purification column (Thermo Scientific) and diluted in hybridization buffer (2× SSC,10%dextran sulfate, 50%formamide, 0.8 mg/mL salmon sperm DNA) to a final volume of 220 µL. We performed FISH according to Grimaud et al. (2005) except that we added hybridization mixture with the probe to the slide before heating. We added a coverslip, sealed it with rubber cement, and heated the slide for 2 min on a 91°C heat block. We obtained embryos (mixed sex) by mating virgin females,aged 3–4 d, of either genotype (1) homozygous MTD or (2) MTD crossed to Bloomington #57008 for clamp RNAi with w1118 males. Embryos were fixed in 3.7% formaldehyde in PBS with an equal volume of n-heptane for 20 min, immunostained using the above antibody concentrations, mounted using Prolong Diamond anti-fade reagent (Thermo Fisher), and imaged on a Zeiss laser scanning 510 or 800 confocal microscope equipped with a 63×/1.4 oil immersion plan apochromat objective and Zen software. We performed polytene chromosome squashes from salivary glands of mixed sex larvae. We passed glands through fix 1 (1.5% formaldehyde, 1% Triton X-100, in 1× PBS) for 1 min, fix 2 (1.5% formaldehyde, 50% glacial acetic acid) for 2 min, and 1:2:3 solution (ratio of lactic acid:water:glacial acetic acid) for 5 min prior to squashing and spreading. Slides were immunostained using the antibody concentrations above and mounted using Prolong Diamond anti-fade reagent (Thermo Fisher), and spreads were imaged on a Zeiss Imager.M1 using a 40×/0.75 plan neofluar objective and AxioVision software.

Western blotting

We conducted Western blotting as in Urban et al. (2017). We collected 2- to 4-h embryos of the relevant genotypes (at least 150 per sample) on grape juice agar plates and washed them briefly with 1× PBS in a cell strainer basket. We dechorionated them for 2 min in 50% bleach

and then washed them with several milliliters of PBS before transferring them to lysis buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1% SDS, 0.5× protease inhibitor). For salivary glands, we dissected glands from third instar larvae (n = 10 per sample) in cold PBS and froze samples in liquid nitrogen. We extracted total protein from samples by homogenizing the samples in cold lysis buffer using a small pestle. We cleared the samples by centrifuging at 14,000g for 10 min at room temperature. To blot for CLAMP and Actin, we ran 20 μg of total protein on a Novex4%-12% Tris-glycine precast gradient gel (Life Technologies). We transferred proteins to PVDF membranes using the iBlot transfer system (Thermo Fisher Scientific) and probed the membranes for CLAMP (rabbit anti-CLAMP, 1:1000; rabbit anti-CLAMP∗, 1:1000) and Actin (mouse anti-Actin, 1:400,000; Millipore) using the Western Breeze kit, following the manufacturer's protocol (Thermo Fisher Scientific).

MNase-seq

We maintained S2 cells in standard Schneider's medium (Gibco) with 10% heat inactivated fetal bovine serum (Thermo Fisher Scientific). We performed RNAi as in Soruco et al. (2013).We performed and analyzed MNase-seq data as in Mieczkowski et al. (2016). We mapped reads to the custom histone locus genome (McKay et al. 2015) using Bowtie aligner with the parameters "-M 5 -k 1 -I 50 -X 500 --solexa-quals --best –chunkmbs 256" (Langmead et al. 2009). We identified genomic positions with abnormally high numbers of mapped reads (Z-score = 7) and discarded tags mapped to such positions. We computed read frequencies in 100-bp nonoverlapping bins and normalized for the library size. We calculated MNase accessibility (MACC) values for each bin by fitting linear regression on the normalized read frequencies computed for each titration point (1.5-, 6.25-, 25-, and 100-U MNase concentrations). We used log scale for the MNase concentrations in the fitting procedure. We applied the GC content

correction to obtain the final accessibility scores (MACC values). The chromatin accessibility data are available at NCBI Gene Expression Omnibus (GEO) with series number GSE99894.

Quantitative real-time PCR

We conducted qRT–PCR as described in Urban et al. (2017) using the embryo RNA obtained for mRNA-seq (below) as well as RNA extracted from unfertilized oocytes laid by unmated mothers and collected 0–2 h after egg lay. We used four biological replicates for each genotype and time point. Primers for histone transcripts H3 and H4 and the normalization gene rp49 are listed in Bulchand et al. (2010), while clamp and pka primers can be found in Urban et al. (2017). We normalized histone transcript abundance against rp49 and clamp transcript abundance against pka. We analyzed data using a Student's t-test, comparing transcript abundance between clamp RNAi embryos or oocytes and matched MTD control embryos or oocytes.

Embryo mRNA-seq

We used embryo RNA collected for qRT–PCR (above). As in Wood et al. (2016), we used 100 ng of total RNA as input for the Ovation Universal RNA-seq kit with Drosophila rRNA depletion module (NuGEN). We sequenced libraries on an Illumina HiSeq 2500 in 1 × 50-bp mode. We used at least four individually isolated biological replicates for each time point and RNAi condition. We mapped reads using TopHat version 2.0.13 with default parameters (Trapnell et al. 2009) and counted fragments mapping to histone gene exons (see Supplemental Table S1 for all FPKM [fragments per kilobase per million mapped fragments] values for histone genes). See Supplemental Table S2 for a list of significantly affected genes. The mRNA-seq data are available at NCBI GEO with series number GSE102922.

Staged embryo ChIP-seq

To obtain female embryos, we mated +; SD72/CyO females to 19-3, yw, Rsp[s] B[s]/Dp(2:y)CB25-4, y+, Rsp[s]B[s]; SPSD/CyO males (both kind gifts from Cynthia Staber) to obtain +/Dp(2:y) CB25-4, y+, Rsp[s]B[s]; SPSD/SD72 males, which we then mated to yw; attP2 PCNA-EGFP females (kind gift from Shelby Blythe). We performed 0- to 4-h timed lays and collected and fixed embryos according to Blythe and Wieschaus (2015). We then handsorted embryos using a Zeiss Discovery.V8 microscope under GFP excitation using an X-CITE 120Q stereo light source. We pooled 200 (NC 11–14) to 400 (NC < 11) embryos and performed ChIP as in Blythe and Wieschaus (2015) using 3 µL of rabbit anti-CLAMP antibody per sample. We synthesized libraries using the NEBNext ChIP-seq kit (New England Biosystems) and sequenced libraries on an Illumina HiSeq 2500 in 2×100-bp mode. We mapped CLAMP ChIP-seq reads to the custom histone locus genome (McKay et al. 2015), allowing only unique alignments by using Bowtie aligner (Langmead et al. 2009).

Analysis of H2a expression from the ectopic array

We isolated total RNA from larvae (n = 10) of the indicated genotypes by flash freezing samples in liquid nitrogen and homogenizing them with a steel bead using a Retsch MM300 TissueLyser Mixer Mill. We then performed phenol/chloroform (Invitrogen) total RNA extractions. We reverse-transcribed 1 µg of total RNA using the SuperScript II kit (Invitrogen) using random primers according to the manufacturer's instructions. We analyzed expression from the transgenic array Histone2a genes via PCR and restriction digestion as in McKay et al. (2015).

RESULTS

CLAMP is recruited to the histone locus via the H3–H4p

While studying CLAMP in the context of male X-chromosome dosage compensation, we noticed distinct CLAMP puncta in the nuclei of early *D. melanogaster* embryos. Costaining revealed that these CLAMP puncta colocalized with markers of the HLB in embryos (Fig. 2.2 A) and cultured cells (Supplemental Fig. S1A) and on the giant salivary gland polytene chromosomes of third instar larvae (Fig. 2.2 B). A GFP-tagged full-length CLAMP also colocalized with HLB markers on salivary gland polytene chromosomes (Supplemental Fig. S1B). Both CLAMP and GAF recognize GA repeats throughout the genome, often at the same loci(Kasinathan et al. 2014; Kuzu et al. 2016). However, we found that GAF was not present at the HLB (Fig. 2.2 C). The endogenous Drosophila histone locus on chromosome 2L contains ∼100 tandem copies of a 5-kb gene cluster(Lifton et al. 1978; McKay et al. 2015), each containing a single copy of the five replication-dependent histone genes (Fig. 2.2D). To determine the exact location of CLAMP binding within the histone locus, we mapped existing *Drosophila* cell culture CLAMP ChIP-seq (chromatin immunoprecipitation [ChIP] combined with high- throughput sequencing) data from our laboratory(Soruco et al. 2013) and GAF ChIP-seq data(Fuda et al. 2015) to a custom genome containing a single copy of the 5-kb gene cluster(McKay et al. 2015). With this approach, the ChIP-seq signal represents an average binding profile across all ∼100 gene clusters. We found that CLAMP localized precisely to the H3–H4p in both male S2 and female Kc cultured cells (Fig. 2.2D), which is the same region of the gene cluster that is minimally sufficient for recruitment of HLB components(Salzler et al. 2013). In marked contrast, GAF did not localize to the histone locus (Fig. 2.2C,D), suggesting

that CLAMP provides a unique function at the histone locus, similar to our results on the dosage-compensated X chromosome(Soruco et al. 2013; Kuzu et al. 2016).

A number of factors are present at the HLB constitutively throughout the cell cycle, while others are present only during S phase, when the histone genes are transcriptionally active(Duronio and Marzluff 2017). The scaffolding protein Mxc is present at the HLB throughout the cell cycle but is phosphorylated only when Cyclin E/Cdk2 is active (e.g., during S phase in cultured cells), creating a phosphoepitope recognized by the MPM-2 antibody(White et al. 2011). Therefore, to characterize whether CLAMP localization to the HLB is cell cycle-dependent, we used the MPM-2 antibody to label S-phase HLBs and the Mxc antibody to label all HLBs. Unlike the MPM-2 epitope, CLAMP was present at the HLBs in all cultured cells (Supplemental Fig. S1A), suggesting that CLAMP localizes to the HLB throughout the cell cycle.

A "proto-HLB" composed of FLASH and Mxc forms before the onset of zygotic histone gene transcription(White et al. 2011; Salzler et al. 2013). However, neither FLASH nor Mxc nor any other previously known component of the HLB has been shown to bind DNA and target the HLB to the histone locus. Our observations that CLAMP localizes to the H3–H4p (Fig. 2.2) and is present at the HLB throughout the cell cycle (Supplemental Fig. S1A) led us to hypothesize that CLAMP may be a factor that is recruited to the histone locus prior to activation of zygotic histone gene expression (i.e., by embryonic nuclear cycle 10) and therefore may be a component of the "proto-HLB." To test this hypothesis, we performed CLAMP ChIP-seq from pools of 200-400 hand-sorted precisely staged embryos(Blythe and Wieschaus 2015). We identified pools of embryos at each of the nuclear cycles 11–14 using a PCNA-EGFP nuclear reporter. We pooled

embryos in cycle 10 and younger to obtain sufficient chromatin from such young embryos for ChIP-seq analysis. We then mapped ChIP-seq reads to the histone gene cluster. CLAMP is present at the embryonic H3– H4p, as observed in cultured cells (Fig. 2.2D), and was present at all assayed nuclear cycles, including by nuclear cycle 10 (Fig. 2.2E). In contrast, other HLB components (i.e., Mxc, FLASH, Mute, and U7 snRNP) are not detectable at nuclear foci prior to cycle 10(White et al. 2007; Terzo et al. 2015). These observations demonstrate that CLAMP is present at the embryonic histone locus prior to zygotic genome activation and suggest that it is recruited to the histone locus before the mature HLB is formed(White et al. 2011; Salzler et al. 2013). Collectively, our observations led us to hypothesize that CLAMP regulates the histone locus during development by recognizing critical cis elements within the H3–H4p.

Figure 2.2

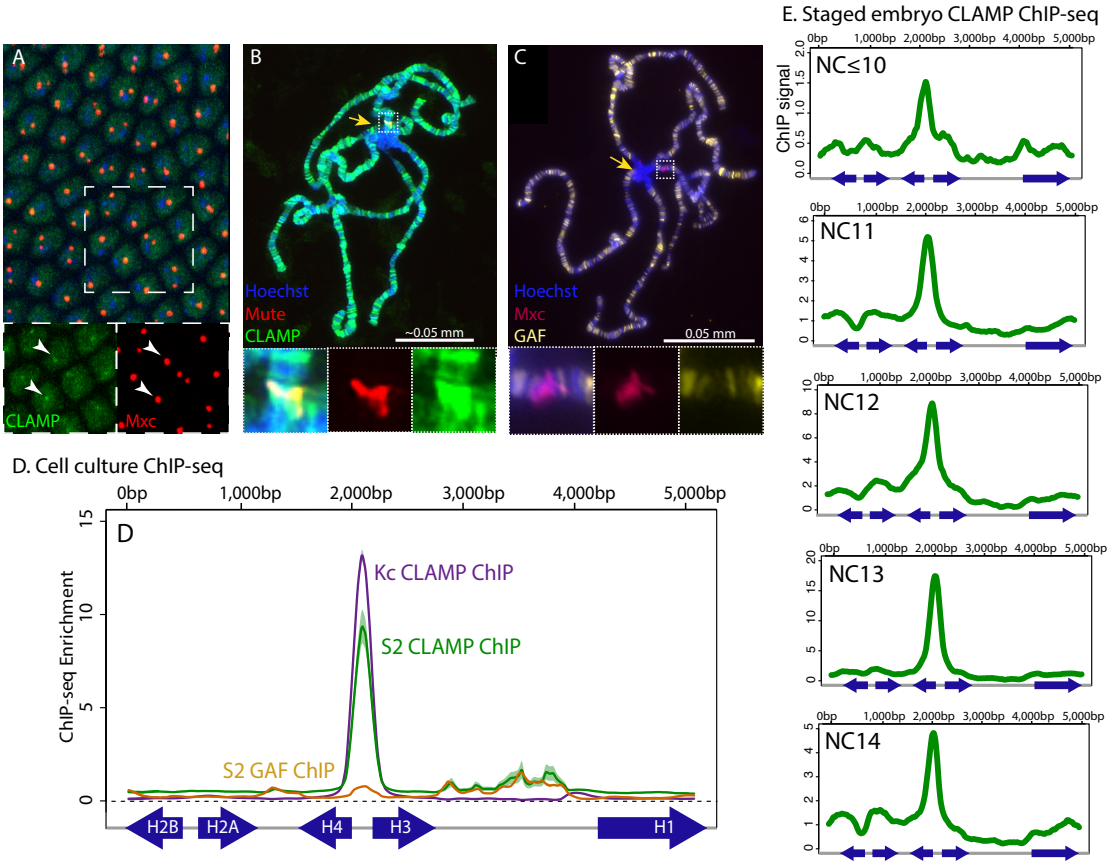


Figure 2.2. CLAMP colocalizes with markers of the HLB. Embryos (A) and third instar larvae salivary gland polytene chromosomes (B,C) immunostained for CLAMP (green), HLB components (Mxc and Mute; red), and GAF (yellow). (A) CLAMP forms distinct puncta in the syncytial nuclei of wild-type Drosophila embryos that colocalize (arrowheads) with Mxc foci. (B) In salivary gland polytene chromosomes, CLAMP colocalizes with Mute at the histone locus near the chromocenter (yellow arrow). (C) GAF, another GA-binding factor, does not colocalize with Mxc. (D) We mapped our previous CLAMP ChIP-seq (chromatin immunoprecipitation [ChIP] combined with highthroughput sequencing) data from cultured S2 (male; green) and Kc (female; purple) cells (Soruco et al. 2013) and existing GAF ChIP-seq (yellow) from cultured S2 cells (Fuda et al. 2015) to the histone gene cluster. Shading represents 95% confidence intervals. ChIP-seq data were normalized to inputs. (E) We performed CLAMP ChIP-seq from precisely staged early embryos and mapped reads to the histone gene cluster. CLAMP is present at the H3–H4p as early as nuclear cycle 10, before zygotic genome activation. We normalized CLAMP ChIP-seq data to ChIP input, as in D.

Conserved GA repeat cis elements in the H3–H4p are required for ectopic HLB formation

We showed previously that a single transgenic copy of the ~300-bp H3–H4p is necessary and sufficient to recruit HLB components to an ectopic locus. We identified two conserved GA repeat motifs within the H3–H4p as potential CLAMP-binding sites (Fig. 2.3A,B; (Salzler et al. 2013)). There is a substantial expansion of one GA repeat in *D. melanogaster* compared with other drosophilids, including closely related species such as *Drosophila simulans* (Fig. 2.3B). Because we reported recently that expanded GA repeats facilitate CLAMP-mediated X-chromosome dosage compensation(Kuzu et al. 2016), we asked whether CLAMP localization to the HLB was specific to *D. melanogaster* by staining polytene chromosomes from *D. simulans* (Supplemental Fig. S2A,B) and *Drosophila virilis* (Fig. 2.3C,D), which diverged from *D. melanogaster* >40 million years ago(Russo, Takezaki, and Nei 1995). The genome of *D. simulans* contains a single histone locus near the chromocenter, while the genome of *D. virilis* contains two histone loci in the middles of chromosome arms(Schienman, Lozovskaya, and Strausbaugh 1998; Berloco et al. 2001). We found that CLAMP is present at the histone locus in both *D. simulans* and *D. virilis*. Similar to *D. melanogaster* (Fig. 2.2C), GAF did not colocalize

with HLB factors in other species. Therefore, CLAMP localization to the histone locus is not specific to *D. melanogaster* and is not only due to the GA repeat expansion in the *D. melanogaster* H3–H4p (Fig. 2.3B).

Our CLAMP ChIP-seq results (Fig. 2.2D,E) and the sequence conservation of the H3–H4p (Fig. 2.3B) led us to hypothesize that the GA repeats may function to promote HLB formation. To identify regions in the H3–H4p that are important for HLB factor recruitment, we constructed four transgenes containing deletions in the ~300-bp H3- H4p. In three of the four constructs, either one or both of the GA repeats are deleted (Supplemental Fig. S3A). We found that HLB factors were efficiently recruited only to the H3–H4p transgene constructs that preserve both GA repeats (Supplemental Fig. S3B–D).

Figure 2.3

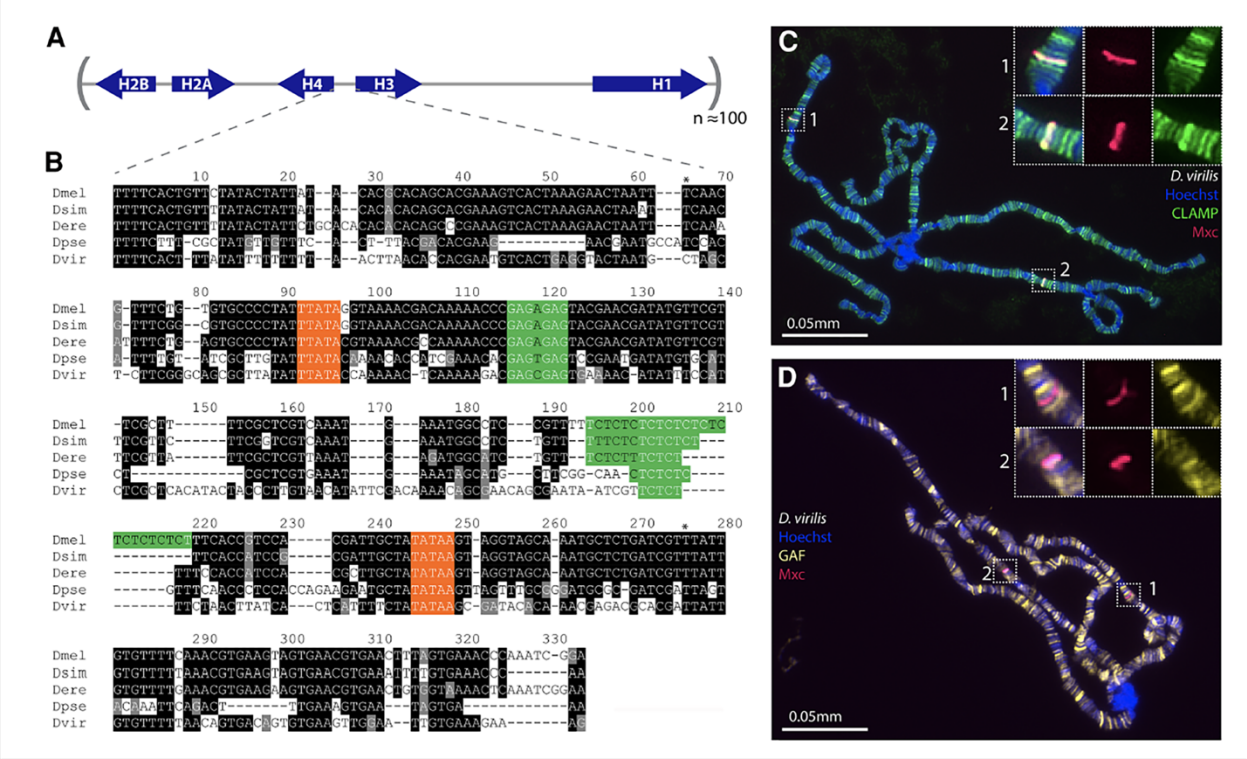


Figure 2.3. Localization of CLAMP at the HLB is conserved across drosophilids. (A) The five replication-dependent histone genes are clustered in a tandemly repeated array of an ~5-kb repeat unit. Approximately 100 repeat units comprise the histone locus on D. melanogaster chromosome 2L. (B) The ~300-bp bidirectional promoter between the Histone3 and Histone4 genes is highly conserved among drosophilids. (Dmel) D. melanogaster; (Dsim) D. simulans; (Dere) Drosophila erecta; (Dpse) Drosophila pseudoobscura; (Dvir) D. virilis. Two GA repeats (green) are highly conserved from D. melanogaster to D. virilis (~40-million-year-ago divergence) (Russo et al. 1995). TATA boxes are highlighted in orange. There is a substantial expansion of one GA repeat in D. melanogaster. Asterisks represent transcription start sites for the H3 and H4 genes. (C,D) We stained polytene chromosomes from D. virilis (which has two histone loci) for Mxc (red), CLAMP (green), and GAF (yellow). Mxc and CLAMP are recruited to both D. virilis histone loci (C), and GAF is not recruited to either histone locus (D).

To test directly whether the GA repeats are required for HLB formation, we used mutational analysis of BAC- based transgenes carrying 12 copies of the full 5-kb histone gene cluster (12xHistone^{WT}). We showed previously that the 12xHistone^{WT} transgene forms a functional ectopic HLB, recruiting all tested HLB factors in the presence of the endogenous histone locus (Fig. 2.4A), and rescues the lethality caused by homozygous deletion of the endogenous histone locus(McKay et al. 2015). The 12xHistone^{WT} transgene also recruits CLAMP (Fig. 2.4A). We therefore generated a mutant 12x array, 12x2xLacO, in which both GA repeats in the H3–H4p are replaced with LacO sequences (Fig. 2.4B; promoter sequence in the Supplemental Material). We integrated this transgene into the *Drosophila* genome at the same site and asked whether it recruits HLB factors in the presence of the endogenous histone locus. The 12x2xLacO transgene fails to recruit CLAMP (Fig. 2.4B), indicating that the GA repeat sequences are necessary for CLAMP binding to the H3–H4p in the presence of the endogenous histone locus. The 12x2xLacO transgene also failed to recruit Mxc, Lsm10, or Mute to polytene chromosome spreads (Fig. 2.4B), consistent with our results from the promoter deletion transgene constructs (Supplemental Fig. S3). In contrast to the 12xHistone^{WT} array(McKay et al.

2015), the 12x2xLacO array in which the GA repeats have been replaced with LacO sequences did not express detectable amounts of Histone2A mRNA (Fig. 2.4E). Thus, the GA repeats in the H3–H4p are key cis elements that promote HLB formation and histone gene expression.

Figure 2.4

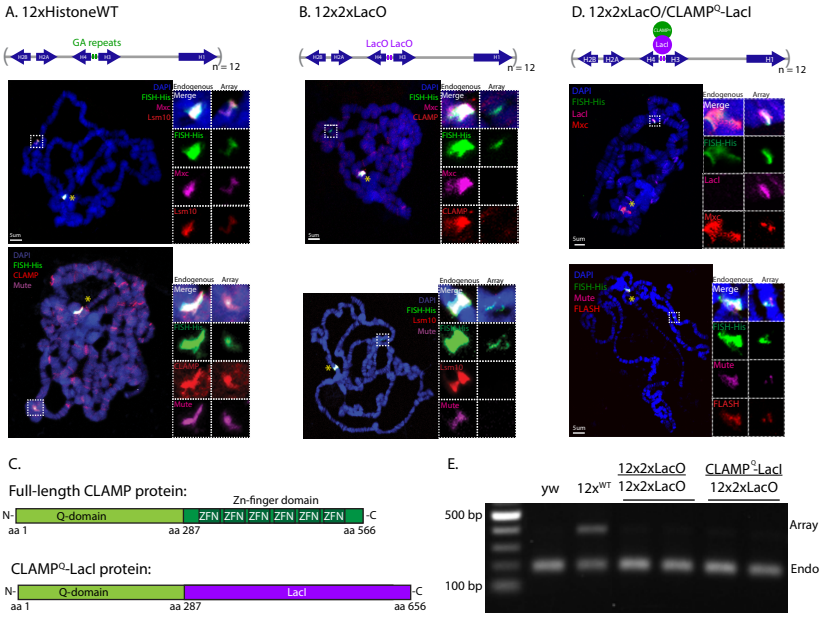


Figure 2.4. **The GA repeats in the H3–H4p are required for HLB formation, and retargeted CLAMP recruits HLB factors.** We integrated transgenes carrying 12 tandem arrays of the wild-type histone gene cluster (A) or a cluster in which the GA repeats are replaced by two LacO sites (B) into the Drosophila genome and scored for recruitment of HLB factors (Mxc, CLAMP, Mute, and Lsm10) by fluorescent in situ hybridization (FISH) against the histone gene cluster (FISH-His; green). The yellow asterisk indicates the endogenous histone locus, while the array locus is boxed. DNA is visualized by DAPI (blue). The sequences of the mutant H3H4p are in the Supplemental Material. (C)We designed the CLAMPQ-LacI fusion protein to include the N-terminal glutamine-rich domain of full-length CLAMP (amino acids 1–287). We replaced the C-terminal zinc finger DNA-binding domain of CLAMP with LacI (purple). (D) We recombined the 12x2xLacO array transgene and the clampq-lacI transgene onto the same chromosome, performed FISH-His (green), and assayed recruitment of CLAMPQ-LacI (detected with anti-LacI antibody; pink) and the HLB components Mxc (red), Mute (pink), and FLASH (red). Yellow asterisks indicate the endogenous histone loci, while array loci are boxed. (E) Expression of Histone2a mRNA from endogenous histone locus and transgenic histone gene cluster arrays. We mutated a site within the transgenic H2a gene (array) to prevent restriction digestion of the H2a cDNA (McKay et al. 2015). Yellow−, white− (yw) larvae do not carry the transgenic locus and produce only histone transcripts from the endogenous locus (cut; Endo), while animals

transgenic for the 12xWT array produce both endogenous and array transcripts (uncut; array). Animals carrying the 12x2xLacO array and animals carrying the 12x2xLacO array and clampq-lacI transgene do not express array transcripts.

CLAMP promotes recruitment of HLB components to an ectopic histone locus

We demonstrated previously that CLAMP recognizes GA-rich sequence elements and promotes recruitment of the MSL complex specifically to the male X chromosome(Soruco et al. 2013; Kuzu et al. 2016). We therefore hypothesized that CLAMP functions similarly at the histone locus, recognizing GA repeat cis elements and promoting the recruitment of HLB-specific factors. We performed a tethering experiment using the 12x2xLacO transgenic array described above (Fig. 2.4B) and a synthetic CLAMP^{Q}-LacI protein in which the CLAMP zinc finger DNA-binding domain(Larschan et al. 2012) is replaced with LacI(Robinett et al. 1996). The resulting transgene, "CLAMP^{Q}-LacI," contains only the N-terminal 287 amino acids of CLAMP, which includes the polyglutamine domain of the CLAMP protein but does not include the DNA-binding domain. We conducted the experiment in this way to avoid the presence of two competing DNA- binding domains on the same protein that could prevent binding to the histone locus (Fig. 2.4C). We expressed the CLAMP^{Q}-LacI transgene using a ubiquitin promoter and measured expression of CLAMP^{Q} -LacI (Supplemental Fig. S4A) by Western blot using protein from third instar larval salivary glands and clamp^{q}-lacI mRNA levels by quantitative RT–PCR (qRT–PCR) (Supplemental Fig. S4B) using RNA from whole larvae. From this analysis, we determined that the CLAMP^{Q}-LacI protein is expressed at higher levels than endogenous CLAMP (Supplemental Fig. S4A).

We performed polytene chromosome immunostaining on animals expressing CLAMP^{Q}-LacI in the presence of the transgenic 12x2xLacO histone array. We found that CLAMP^{Q}-LacI as well as Mxc, FLASH, and Mute colocalized with the fluorescent in situ hybridization (FISH) signal against the histone gene repeat when paired with the 12x2xLacO histone array (Fig. 2.4D). In contrast, in the absence of CLAMP^{Q}-LacI, the 12x2xLacO array does not recruit HLB factors (Fig. 2.4B). Therefore, CLAMP^{Q}- LacI binds the 12x2xLacO array and promotes the recruitment of HLB factors. CLAMP^{Q}-LacI also localized to several additional genomic locations (Fig. 2.4D), possibly through dimerization with endogenous CLAMP. However, HLB factors were recruited only to the 12x2xLacO locus after CLAMP^{Q}-LacI expression. Thus, like endogenous CLAMP, CLAMP^{Q}-LacI is not sufficient to recruit HLB factors to nonhistone loci, suggesting that the H3-H4p may contain other important cis elements and/or recruit additional critical components. Therefore, promoting HLB factor recruitment to the histone locus is a context-specific function of the ubiquitously expressed CLAMP protein, similar to its function in promoting the recruitment of the MSL complex to the dosage-compensated X chromosome.

Although multiple HLB factors were recruited to the 12x2xLacO histone array in the presence of CLAMP^{Q}- LacI, we did not detect any histone mRNA expressed from the array (Fig. 2.4E). Thus, HLB factor recruitment can be uncoupled from histone locus transcription, as observed previously in the formation of a "proto-HLB" containing Mxc and FLASH(Salzler et al. 2013). There are several possible explanations for why CLAMP^{Q}- LacI can recruit HLB factors but not promote transcriptional activation. First, only low levels of HLB factors were recruited to the 12x2xLacO histone array in the presence of CLAMP^{Q}-LacI (Fig. 2.4D),

compared with the levels recruited to the 12xHistone^{WT} array in the presence of endogenous wild-type CLAMP (Fig. 2.4A). Therefore, the stoichiometry of CLAMP^{Q} -LacI compared with each 12x2xLacO sequence may be insufficient to support transcription. Second, a specific conformation of the endogenous CLAMP protein or the zinc finger domain of CLAMP, which is replaced by LacI in the CLAMP^{Q}-LacI construct (Fig. 2.4C), may be required for transcriptional activation of the histone genes but not for recruitment of HLB factors. Alternatively, wild-type CLAMP at the endogenous histone locus may recruit critical limiting factors much more efficiently than ectopically localized CLAMP^{Q}-LacI. Together, our data indicate that the N-terminal 287 amino acids of CLAMP are sufficient for promoting the recruitment of HLB factors to an ectopic chromosomal location but not for transcriptional activation.

CLAMP regulates histone locus chromatin and histone gene expression

Because CLAMP localizes to the histone locus and binds directly to the essential GA repeat cis elements, we hypothesized that CLAMP regulates histone locus chromatin and histone gene transcription. All five histone transcripts and the clamp transcript are expressed in the oocyte and maternally deposited in 0- to 2-h embryos. Mature HLBs are formed by cycle 11, and the zygotic histone locus is activated by zygotic genome activation (~2 h after egg lay)(White et al. 2007). To determine how CLAMP functions as the HLB forms and the histone locus becomes transcriptionally activated in the early embryo, we depleted maternally deposited CLAMP protein and mRNA by RNAi using a UAS-driven clamp-specific shRNA(Ni et al. 2008) and a strong GAL4 driver (maternal triple driver [MTD]) that is expressed in the female germ-line(Staller et al. 2013). As controls, we also analyzed matched MTD controls. We performed qRT–PCR for histone mRNA levels using primers specific to the H3 and H4

transcripts(Bulchand et al. 2010). CLAMP mRNA and protein are nearly completely depleted in 0- to 2-h and 2- to 4-h embryos derived from clamp RNAi mothers (Figs. 2.5 A, 2.6A). We found that in unfertilized eggs and 0- to 2-h embryos laid by clamp RNAi mothers, both H3 and H4 transcript levels were significantly decreased (Fig. 2.5A), indicating that CLAMP depletion results in a reduction in the amount of histone mRNA deposited in the egg. Histone transcript levels began to recover in older embryos, likely due to zygotic genome activation of both the embryonic histone locus and the clamp locus. Nearly 100% (99.87%) of clamp RNAi embryos do not hatch, demonstrating that maternally deposited CLAMP is critical for early development.

Figure 2.5

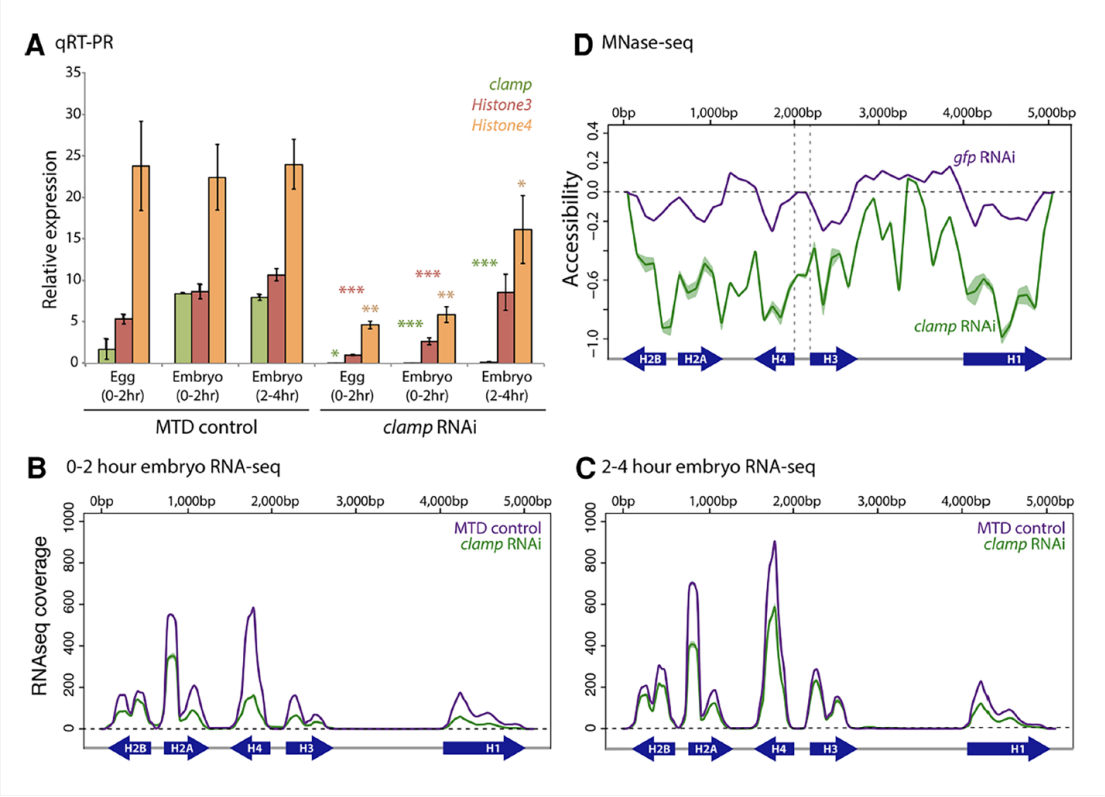


Figure 2.5. CLAMP regulates histone gene transcription and histone locus chromatin accessibility. (A–C) We performed qRT– PCR and mRNA sequencing (mRNA-seq) using RNA from unfertilized eggs laid by virgin mothers, 0- to 2-h fertilized embryos, and 2- to 4-h fertilized

embryos laid by MTD control mothers or clamp RNAi mothers. (A) clamp (green), Histone3 (red), and Histone4 (orange) transcripts are significantly reduced at all stages (with the exception of Histone3 in 2- to 4-h embryos) after clamp RNAi compared with MTD controls, although transcripts begin to recover in 2- to 4-h embryos after zygotic genome activation. Error bars represent ±one standard deviation from the mean. clamp expression was normalized to pka, and Histone3 and Histone4 expression was normalized to rp49. (∗) $P < 0.05$; (∗∗) $P < 0.001$; (∗∗∗) $P < 0.0001$. Western blots for CLAMP protein from 2- to 4-h embryos are shown in Figure 5A. (B) Maternal deposition of all five histone mRNAs is significantly reduced in 0- to 2-h embryos from clamp RNAi mothers (green) compared with control MTD mothers (purple), as assayed by mRNA-seq. (C) Histone transcript levels begin to recover in 2- to 4-h embryos. See Supplemental Table S1 for raw FPKM (fragments per kilobase per million mapped fragments) data and P-values. (D)We analyzed MNase-seq (micrococcal nuclease [MNase] digestion followed by high-throughput sequencing) data from cultured S2 cells treated with clamp RNAi (green) or gfp RNAi (control; purple). Accessibility scores >0 were assigned to chromatin that is more open compared with the rest of the genome, while scores <0 were assigned to chromatin that is relatively closed. (B–D) Shading represents 95% confidence intervals.

We measured transcript accumulation from all histone genes by performing mRNA sequencing (mRNA-seq) from the same embryo time points and RNAi conditions that we used in our qRT–PCR assay (Fig. 2.5A). We found that maternal clamp RNAi resulted in decreased expression of all histone genes compared with MTD controls (Fig. 2.5 B,C; Supplemental Table S1), including H1, H2A, and H2B, which are not adjacent to the site of CLAMP binding within the H3–H4p (Fig. 2.2D,E). To determine whether this reduction of histone transcripts might be an indirect effect of CLAMP regulating the expression of known HLB components, we

compared the levels of mRNAs encoding known HLB components, including FLASH, Mxc, Mute, Lsm10, and Lsm11 (Duronio and Marzluff 2017), from 0- to 2-h and 2- to 4-h clamp RNAi embryos with those from control MTD embryos. The transcript levels of the known HLB factors were not affected by clamp RNAi in the early embryo (Supplemental Table S2). Together, these observations suggest that the effects on histone transcript levels that we observed in the early developing embryo after CLAMP depletion (Fig. 2.5A–C) are not due to misregulation of genes encoding other HLB components but a specific effect of CLAMP depletion on histone mRNA deposition in the egg.

We next sought to understand how CLAMP might regulate histone gene expression. To determine whether CLAMP promotes chromatin accessibility at the histone locus, we performed clamp RNAi in cultured male S2 cells and measured chromatin accessibility using a previously described micrococcal nuclease (MNase) digestion followed by high-throughput sequencing (MNase-seq) titration approach developed by our group(Mieczkowski et al. 2016). Accessible chromatin is digested even under dilute MNase concentrations, while inaccessible chromatin is protected until it is subjected to higher concentrations. Therefore, sequencing libraries generated from chromatin subjected to a low concentration of MNase are enriched for accessible regions, while libraries from chromatin digested with a higher concentration of enzyme are enriched for more inaccessible regions(Mieczkowski et al. 2016).

Using our chromatin accessibility data, we mapped reads to a single copy of the histone gene array(McKay et al. 2015) and calculated chromatin accessibility as described in Mieczkowski et al. (2016). In control S2 cells subjected to gfp control RNAi, the most accessible regions of the histone gene cluster included promoter and intergenic regions (Fig. 2.5D). In cells subjected to clamp RNAi, the entire histone gene cluster decreased in accessibility. Overall, our

data support the requirement for CLAMP recruitment to the histone locus for both promoting chromatin accessibility across the entire histone locus and activating gene expression of all replication-dependent histone genes (Fig. 2.5A–D).

Figure 2.6

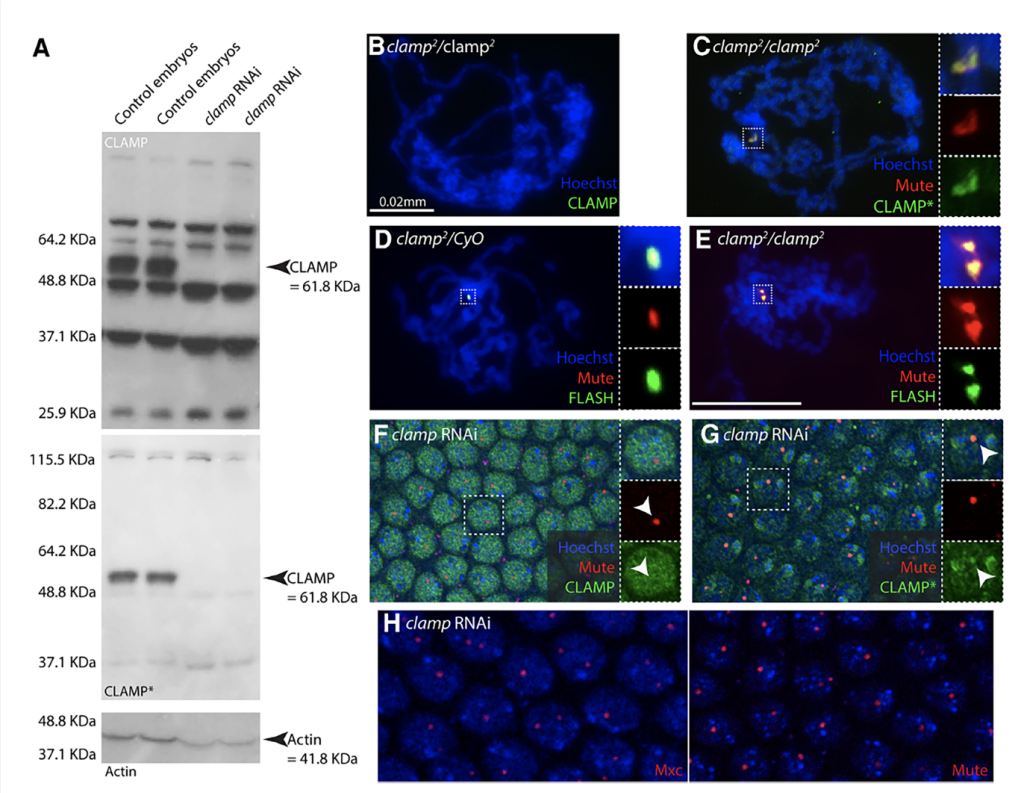


Figure 2.6. CLAMP remains at the histone locus in CLAMP-depleted larvae and embryos. (A) Western blots demonstrating the antibody specificity and efficacy of clamp RNAi in embryos (2–4 h). Control embryos (those expressing the MTDwithout clamp RNAi) have abundant CLAMP protein (61.8 kDa), while CLAMP is undetectable in embryos subjected to clamp RNAi. We developed two custom antibodies against CLAMP. The CLAMP antibody (top) (Larschan et al. 2012; Urban et al. 2017) cross-reacts with several bands by Western blot, while the CLAMP∗ antibody (bottom) has lower cross-reactivity. Note that these Westerns also demonstrate the efficacy of clamp RNAi in 2- to 4-h embryos. We stained for Actin as a loading control. (B–E) Larval salivary gland polytene chromosome spreads from salivary glands from homozygous (clamp2/clamp2; B,C,E) and heterozygous (clamp2/CyO; D) third instar larvae stained for CLAMP (CLAMP [B] and CLAMP∗ [C]; both green) and HLB components Mute (red; C–E) and FLASH (green; D,E). HLBs are formed in both clamp2/ CyO controls (D) and clamp2/clamp2 nulls (C,E). (F–H) clamp-depleted embryos immunostained for CLAMP (CLAMP [F] and CLAMP∗ [G]; both green) and the HLB factors Mute (red) and Mxc (red).

CLAMP is specifically retained at the HLB in clamp nulls

The ability of CLAMP to promote HLB factor recruitment to the transgenic array locus (Fig. 2.4D) and regulate chromatin organization at the histone locus (Fig. 2.5D) led us to test whether removing CLAMP resulted in loss of endogenous HLB formation in clamp-null larvae and in embryos in which clamp is depleted by RNAi. While almost all embryos laid by mothers with RNAi-reduced clamp levels are inviable, ~30% of zygotically mutant animals homozygous for the clamp^{2}-null allele survive until the third instar larval stage(Urban et al. 2017). These data indicate that survival of clamp^{2} larvae is likely due to maternal deposition of wild-type clamp mRNA or protein. We reported previously that clamp^{2}-null animals produce undetectable amounts of CLAMP protein by both Western blot of salivary gland tissue and polytene chromosome immunostaining(Urban et al. 2017). However, our previously characterized affinity-purified anti-CLAMP anti- body detected many cross-reacting proteins by Western blot (Fig. 2.6A), and we suspected that it might not be sufficiently sensitive to detect small amounts of CLAMP protein in clamp^{2}-null larvae. Therefore, we developed a new affinity-purified anti-CLAMP antibody ("CLAMP* ")— produced using a different approach (see the Materials and Methods)—that is more specific and does not cross-react with other *Drosophila* proteins (Fig. 2.6A).

Using both anti-CLAMP antibodies, we analyzed CLAMP localization and HLB formation on clamp^{2}-null polytene chromosomes from larval salivary glands. In addition, we analyzed early embryos after RNAi depletion of maternally deposited CLAMP protein and mRNA (Figs. 2.5A, 2.6A). While these embryos die, it is still possible to analyze CLAMP localization and HLB formation during the early syncytial stages (0–2 h) of embryonic

development. Using the CLAMP antibody, we observed loss of CLAMP from all loci on polytene chromosomes from clamp^{2} nulls, as observed previously (Fig. 2.6B). There were no CLAMP foci in early embryos after CLAMP RNAi, although Mute foci were readily detected (Fig. 2.6F). However, using the CLAMP* antibody, we detected a modest amount of CLAMP immunostaining, which colocalizes specifically and solely with HLB factors on clamp^{2}-null mutant polytenes (Fig. 2.6C) and in embryos after clamp RNAi (Fig. 2.6G). Therefore, the CLAMP* antibody is more sensitive than our previously published CLAMP antibody for immunofluorescence. Multiple HLB factors continue to colocalize on clamp^{2}-null polytene chromosomes, similar to control chromosomes where abundant CLAMP is present (Fig. 2.6 D,E), and in clamp-depleted embryos (Fig. 2.6H). Because zygotic CLAMP protein is not produced in clamp^{2} homozygous animals(Urban et al. 2017), we conclude that maternally deposited CLAMP from clamp^{2} heterozygous mothers persists at the histone locus throughout development of the larval salivary gland. Furthermore, the strong MTD does not deplete all maternally deposited clamp transcript or protein, and therefore a small amount of CLAMP remains at embryonic HLBs. Some of the ~200 GA repeat CLAMP-binding sites within the histone gene array are likely capable of recruiting the small amount of maternal CLAMP that remains in clamp^{2}-null third instar larvae and early clamp-depleted embryos.

GAF localizes to the HLB when CLAMP is depleted

Since only trace amounts of maternally deposited CLAMP are present in HLBs in clamp^{2}-null larvae and embryos from clamp RNAi mothers (Fig. 2.6C,G), we tested whether CLAMP depletion results in novel recruitment of the GAF protein to some H3–H4ps. Consistent with this possibility, we found that GAF colocalized with HLB markers on polytene

chromosomes from clamp^{2} homozygous larvae (Fig. 2.7B) and in syncytial embryos after MTD-driven clamp RNAi (Fig. 2.7D) even though GAF was absent from the HLBs when CLAMP is present (Figs. 2.2C, 2.7A,C). Previous in vitro experiments suggested that GAF is able to bind the GA-rich cis elements in the H3–H4p(Gilmour, Thomas, and Elgin 1989). However, consistent with our current observations, later in vivo studies suggested that GAF is not present at the histone locus(O'Brien et al. 1995) (Bhat et al. 1996). Our results suggest that the 61- kDa CLAMP protein(Urban et al. 2017) (Urban et al. 2017) is the likely ~66-kDa protein bound to the H3–H4p in vivo that was identified >25 years ago(Gilmour, Thomas, and Elgin 1989; Weber and Gilmour 1995). However, in agreement with previous in vitro observations (Weber and Gilmour 1995), GAF localizes to the histone locus when CLAMP is depleted (Fig. 2.7B,D), suggesting that it can bind to the H3–H4p when there is not sufficient CLAMP to occupy all GA repeat-binding sites. There are several models that could explain the relationship between CLAMP and GAF at the HLB. For example, residual amounts of CLAMP (Fig. 2.6) may be sufficient to open the chromatin and initiate HLB formation but not stimulate transcription. In this situation, although GAF can now bind unoccupied GA repeats, it is not able to stimulate histone gene transcription to the level that CLAMP does (Fig. 2.5A–C). A second possibility is that GAF can open chromatin to promote HLB factor recruitment but cannot stimulate transcription at the histone locus to the same level as CLAMP because it cannot interact with the same cofactors as CLAMP. We discuss possible functional relationships between GAF and CLAMP further in the Discussion.

Figure 2.7

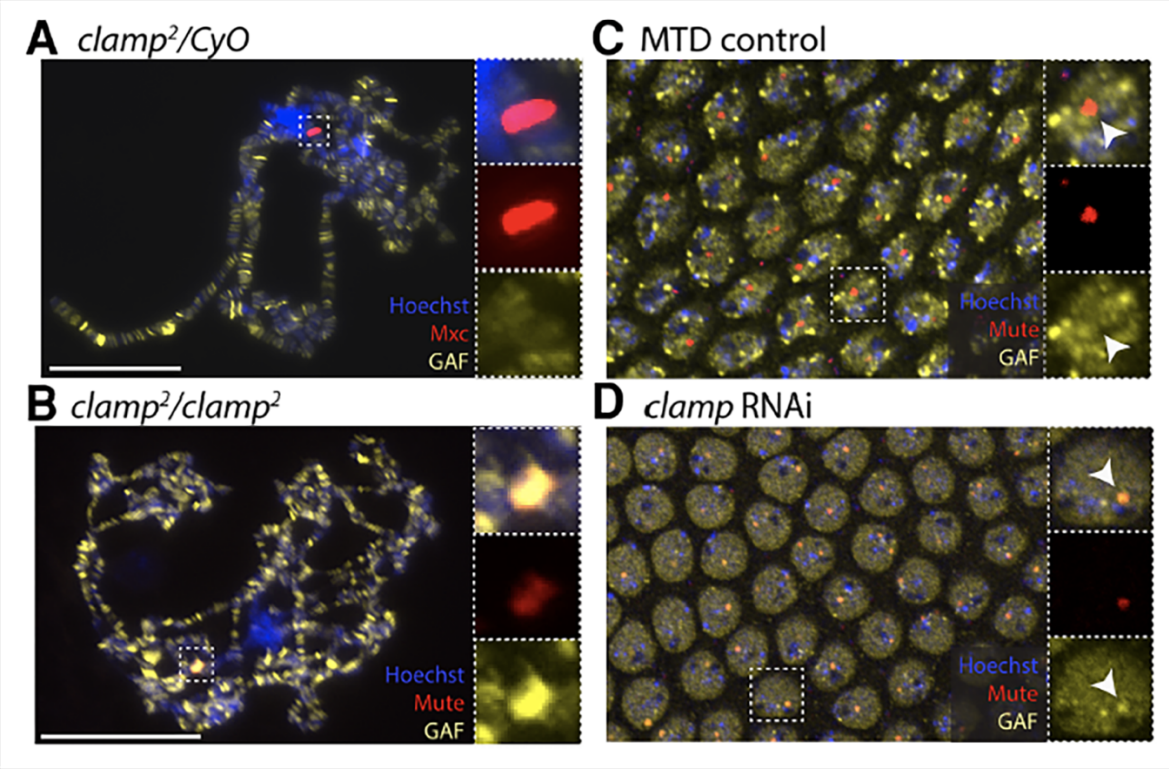


Figure 2.7. GAF localizes to the HLB when CLAMP is depleted. GAF (yellow) does not localize to the HLB (Mxc or Mute; red) in polytene chromosomes from clamp2/CyO heterozygous larvae (A) or in syncytial embryos laid by MTD control mothers (C). However, GAF colocalizes with HLB markers in polytene chromosomes from clamp2 homozygote larvae (B) and in nuclei from syncytial embryos laid by MTD-GAL4/UAS-clamp shRNA mothers (D).

DISCUSSION

NBs provide a means to coordinate gene expression at specific sites in the nucleus. For example, the HLB coordinates histone gene expression to maintain correct nucleosomal stoichiometry. Important open questions include the following: (1) How does the HLB form specifically at the histone locus? (2) How are the histone genes coordinately regulated? Here, we show that two GA repeat arrays in the H3–H4p direct HLB assembly and that a GA repeat-binding factor, CLAMP, localizes to this region early in development, where it regulates the

histone locus by promoting enhanced chromatin accessibility and expression of all five histone genes.

Regulation of the histone locus

The mature Drosophila HLB is formed by cycle 11 during early embryogenesis, prior to widespread zygotic genome activation (White et al. 2011). Activation of histone genes is critical for organismal development and viability(Saget et al. 1998; Godfrey et al. 2006; Godfrey et al. 2009; White et al. 2007; Gunesdogan, Jackle, and Herzig 2010). CLAMP is present at the histone locus when the proto-HLB containing Mxc and FLASH is first detectable (Fig. 2.2E), prior to zygotic histone gene activation (White et al. 2007). As a DNA-binding transcription factor that directly recognizes the GA repeats that promote HLB formation in the presence of the endogenous locus (Fig. 2.2D,E), CLAMP represents a potential pioneer factor involved in promoting HLB formation and activation of histone gene transcription. Furthermore, CLAMP facilitates chromatin opening across the entire histone locus repeat (Fig. 2.5D), suggesting that it could recruit additional factors, such as general transcription factors or cofactors necessary for histone gene regulation.

It is likely that CLAMP affects chromatin accessibility at the histone locus through recruitment of a chromatin-remodeling factor. For example, the similar GAF protein recruits the NURF301 chromatin remodeler(Tsukiyama and Wu 1995). In addition, CLAMP may control histone gene transcription (Fig. 2.5A–C) through modulating chromatin accessibility, or changes in accessibility may be caused by CLAMP directly regulating histone transcription. Retargeting the polyglutamine domain of CLAMP to a synthetic histone array transgene rescues recruitment of Mxc (Fig. 2.4E), the core HLB scaffolding protein(Terzo et al. 2015), and several other HLB components. However, $CLAMP^{Q}$-LacI is not sufficient to activate transcription (Fig. 2.4E). It is

possible that transcriptional activation re- quires additional domains of CLAMP that were removed in our CLAMP^{Q}-LacI fusion protein, a particular conformation of the wild-type CLAMP protein that is not attainable by CLAMP^{Q}-LacI, or higher CLAMP occupancy than the LacO/LacI system allows. Overall, we demonstrate that a reduction in wild-type CLAMP levels leads to decreased chromatin accessibility and histone gene expression at the endogenous histone locus. Furthermore, artificially tethering CLAMP^{Q}-LacI to a synthetic histone gene array recruits HLB-specific factors but does not stimulate histone gene expression.

Interestingly, previous work demonstrated that additional HLB components, including Mute and U7 snRNP, are recruited to the HLB only when transcription is initiated(Salzler et al. 2013) and are present at the mature HLB by nuclear cycle 11, when the histone genes are expressed(White et al. 2007; White et al. 2011). However, in the present study, we demonstrate that recruitment of Mute can be uncoupled from histone locus transcription (Fig. 2.4E), perhaps capturing an intermediate stage in HLB formation.

Conservation of HLB formation

Factors that regulate histone gene expression are well conserved, but current evidence suggests that cis elements that target these factors are less conserved across species. For example, critical components of the HLB that do not directly interact with DNA are conserved across metazoans, including Mxc/NPAT, Mute/YY1-associated protein, FLASH, U7 snRNP, and Coilin(Duronio and Marzluff 2017). In contrast, CLAMP is highly conserved across *Diptera(Kuzu et al. 2016)*, but we did not identify a CLAMP ortholog outside of insects. CLAMP may function together with an early acting *Drosophila*-specific transcription factor, such as the master zygotic genome activator Zelda, which, although conserved among

drosophilids, also has no identified mammalian ortholog(Paris et al. 2013). The mammalian histone genes are more dispersed and there are no repeated units, making it possible that mammalian cis elements may be within poorly conserved noncoding sequences. Therefore, the plasticity of the components that link highly conserved HLB factors to the histone locus—and the flexibility of their low-complexity domains—may allow for evolutionary variation among critical cis-acting sequences that specify HLB formation. CLAMP binding to GA repeats likely provides this function in Drosophila, while other DNA-binding proteins and cis elements may nucleate the HLB in other species.

A single transcription factor can mediate multiple domains of coordinated gene activation

We originally identified CLAMP as an essential factor(Urban et al. 2017) that is required to nucleate the formation of a domain of coordinated gene activation during male X-chromosome dosage compensation(Larschan et al. 2012; Soruco et al. 2013). GA repeat expansion and elevated GA repeat density on the X chromosome evolved to increase CLAMP occupancy and promote MSL recognition(Kuzu et al. 2016). Here we show that CLAMP is also recruited to the histone locus via GA repeats in the H3–H4p (Figs. 2.2 D,E, 2.2B), where it promotes the formation of another domain of coordinated gene activation. In contrast to CLAMP, GAF is not required for male X-chromo- some dosage compensation(Greenberg, Yanowitz, and Schedl 2004) and is dramatically depleted on the X chromosome compared with autosomes(Soruco et al. 2013). GAF is also not pre- sent at the histone locus in wild-type situations (Fig. 2.2C, D). Therefore, we identified the enrichment of CLAMP relative to GAF as a common feature of two different chromatin domains of enhanced chromatin accessibility and active transcription.

Although CLAMP regulates two very different active chromatin domains to which different specific cofactors are recruited, CLAMP is also a transcription factor that binds to thousands of loci throughout the genome. The mechanism by which a single protein fulfills multiple context-specific roles within the genome is not well understood in any system. It is likely that synergy with additional cofactors, such as the MSL complex on the dosage-compensated X chromosome(Soruco et al. 2013) and Mxc at the histone locus, are critical for this process.

In contrast to the histone locus, where HLB-specific factors are conserved and cis sequences are not, the MSL dosage compensation complex components are very poorly conserved even within drosophilids, compared with CLAMP. However, X-enriched GA-rich cis elements are well conserved across insect species(Kuzu et al. 2016). Because CLAMP is less conserved across species compared with HLB-specific factors, it is possible that it provides the plasticity required to couple conserved regulators to rapidly evolving cis elements at the histone locus. In contrast, on the X chromosome, CLAMP functions as a relatively well-conserved factor (compared with the poorly conserved MSL complex) that couples conserved cis elements to diverged cofactors. It is possible that the ancient evolutionary role of CLAMP is as a GA-binding transcription factor, while the context-specific roles of CLAMP at the histone locus and dosage-compensated X chromosome evolved more recently, after GA-rich cis elements became enriched at these locations to promote domains of coordinated gene activation. However, significant further study is required to fully understand how a single protein promotes the formation of multiple active chromatin domains.

The relationship between GA-binding factors at the histone locus

GAF is a ubiquitously expressed transcription factor that binds GA repeats genome-wide and functions as a transcriptional regulator(Horard et al. 2000; Fuda et al. 2015), chromatin modifier(Tsukiyama and Wu 1995), and insulator(Ohtsuki and Levine 1998), depending on the context. Although GAF is excluded from the histone locus in wild-type situations, it localizes to the histone locus when CLAMP is depleted (Fig. 2.7).

There are several possible models for the relationship between CLAMP and GAF at the HLB: (1) CLAMP and GAF could be partially redundant such that GAF can compensate for some functions normally accomplished by CLAMP, including opening chromatin to minimally recruit HLB factors. However, GAF is not sufficient to fully rescue histone transcription levels when CLAMP is depleted (Fig. 2.5A–C). (2) CLAMP and GAF could have an antagonistic relationship at the histone locus such that a major function for CLAMP could be to exclude GAF. For example, GAF mislocalization to the histone locus may close chromatin and repress histone gene expression through its well-documented relationship with the repressive Polycomb complex(Horard et al. 2000). (3) Both CLAMP and GAF contain low-complexity glutamine-rich domains(Tariq et al. 2013) that are thought to modulate transcriptional activity(Gemayel et al. 2015) and are also frequently found in a class of proteins thought to undergo biophysical phase transitions to drive NB formation(Zhu and Brangwynne 2015). Therefore, the relative occupancy of CLAMP compared with GAF within a domain may promote or hinder NB formation. (4) GAF binds to unoccupied histone locus GA repeats opportunistically and does not function redundantly or antagonistically with CLAMP at the histone locus. In this model, CLAMP depletion drives the changes that we observed in histone locus regulation (Fig. 2.5).

Since we were not able to completely remove CLAMP even in the salivary glands of clamp-null mutant larvae, we conclude that CLAMP regulates the histone locus (Fig. 2.6C,G) but cannot conclude that CLAMP is essential for endogenous HLB formation. In addition, in agreement with previous studies on GAF autoregulation(Bernues, Pineyro, and Kosoy 2007), we found that the GAF-encoding mRNA tri-thorax-like is resistant to depletion by embryonic RNAi using the shRNA system. Therefore, further experimentation is required to determine the functional relationship between CLAMP and GAF at the HLB. Here, we determined that the same critical cis elements (GA repeats) and trans-acting factor (CLAMP) and the absence of a second factor that binds similar cis elements (GAF) promote the formation of two different domains of coordinated gene activation. Furthermore, both at the histone locus and on the dosage-compensated male X chromosome, CLAMP is required to promote a domain of enhanced chromatin accessibility that extends well beyond its binding sites. Overall, our study provides new insight into the common regulatory mechanisms shared by two different domains of coordinated gene activation.

ACKNOWLEDGMENTS

We thank Victoria Cabot for initial observations; Shelby Blythe for the PCNA-EGFP fly stock and early embryo ChIP-seq protocol; Cindi Staber for SD and Rsp fly stocks; members of the Anne Hart and Casey Dunn laboratories for use of their fluorescent dissecting microscopes; Judith Bender, Arjun Ray, Jason Wood, and Steve Criscione for advice on experimental analyses that did not end up in the final manuscript; Daniel McKay for the custom histone array genome; Deirdre Tatomer for several published fly stocks; and Geoff Williams for microscopy advice.We thank Joseph Gall for an anti-Coilin antibody, Giacomo Cavalli for the anti-GAF antibody, Jacob Bliss for experimental conversations, and Brian Jones and Judith Bender for

critical comments on the manuscript. We thank the Drosophila Genomics Resource Center, the Transgenic RNAi Project (TRiP), and the Bloomington Stock Center for fly lines. Finally, we thank anonymous reviewers for their constructive feedback. L.E.R. is supported by National Institutes of Health (NIH) grant F32 GM109663, and K.A.B. was supported by NIH training grant T32 CA009156. All experiments were supported by NIH grants GM098461 (E.N.L.) and GM058921 (W.F.M and R.J.D.), American Cancer Society Research Scholar grant 123682-RSG-13-040-01-DMC (E.N.L.), and a Pew Biomedical Scholars program grant (E.N.L.). L.E.R., K.P.K, K.A.B., E.N.L., R.J.D., and W.F.M. conceived the study. L.E.R., K.P.K., K.A.B., J.A.U., S.B., and A.Z. performed the investigations. L.E.R., G.K., and M.Y.T. performed the formal analysis. L.E.R. wrote the original draft of the manuscript, and L.E.R., E.N.L., R.J.D., and W.F.M. reviewed and edited the manuscript. L.E.R., K.P.K., and K.A.B. performed visualization. M.Y.T. and W.T.J. supervised the study. L.E.R., K.A.B.,E.N.L., R.J.D., and W.F.M. acquired the funding.

CHAPTER 3: HISTONE LOCUS BODY FORMATION: DIFFERENT WAYS TO A COMMON END

INTRODUCTION

An important organizing principle in cells is the use of membraneless compartments to spatially and temporally regulate diverse biological processes (Mitrea and Kriwacki 2016). Numerous membraneless compartments have been identified in both the nucleus (e.g. nucleoli, Cajal bodies, histone locus bodies) and the cytoplasm (e.g. P-bodies, stress granules, germ granules) and are collectively referred to as biomolecular condensates (Banani et al. 2017). These structures can be observed in the light microscope and are defined by the their ability to concentrate factors involved in specific processes(Banani et al. 2017)·(Dundr and Misteli 2010). Biomolecular condensates found in the nucleus are often referred to as nuclear bodies (NBs), and represent one of the major organizing features of the nucleus. By concentrating factors, NBs are postulated to increase biochemical reaction efficiency and thus provide a benefit to the cell (Sawyer and Dundr 2016; Stroberg and Schnell 2018). Although evidence for this hypothesis exists (Strzelecka et al. 2010; Novotny et al. 2011; Walker, Tian, and Matera 2009; Tatomer et al. 2016a), the disruption of a NB doesn't always result in an obvious impact to the nuclear process with which the body is associated (Liu et al. 2009; Collier et al. 2006; Deryusheva and Gall 2009). Therefore, ascribing functional roles for NBs has remained somewhat problematic.

The Histone Locus Body (HLB), is a conserved NB that assembles at replication-dependent (RD) histone genes and contains factors required for RD histone mRNA biogenesis. The RD histone mRNAs are the only eukaryotic non-polyadenylated mRNAs that have been

identified, even after multiple rounds of deep sequencing (Marzluff and Koreski 2017). The unique stem loop at the 3'end of RD histone mRNAs results from a processing reaction requiring a specialized suite of factors, some of which are constitutively localized in the HLB (Duronio and Marzluff 2017). With a well characterized set of factors involved in producing a unique class of cell-cycle regulated mRNAs, the HLB provides a powerful system to study how NBs form and function. We previously demonstrated that concentrating factors (e.g. FLASH and U7 snRNP) in the *Drosophila melanogaster* HLB is critical for efficient histone pre-mRNA processing (Tatomer et al. 2016a; Wagner et al. 2007). However, a full determination of how the HLB participates in histone mRNA biosynthesis requires a detailed molecular understanding of HLB assembly.

Prior studies of NBs have provided several important assembly concepts that are applicable to the HLB. Many NB components have an intrinsic ability to self-associate, an observation leading to two models of NB assembly: (1) interactions among NB components occur stochastically wherein individual factors can be recruited to the body in any order or (2) an ordered or hierarchical assembly pathway wherein the recruitment of some components is predicated on prior recruitment of others (Dundr and Misteli 2010; Rajendra, Praveen, and Matera 2010). The HLB appears to employ a hybrid version of these two extreme possibilities. Genetic loss of function experiments suggest a partially ordered assembly pathway with some components being required for subsequent recruitment of others (White et al. 2011), while tethering experiments indicate that ectopic HLB formation may be induced by several different HLB components, supporting a stochastic model of assembly (Shevtsov and Dundr 2011).

Irrespective of the manner of factor recruitment, the initiation event in self-organizing NB assembly is a key, but incompletely understood, step in the process. A prevalent model

postulates a “seeding” event that initiates the nucleation of critical components that form a platform for further recruitment of other components (Shevtsov and Dundr 2011; Dundr 2011; Falahati et al. 2016; Altmeyer et al. 2015; Gomes and Shorter 2018; Staněk and Fox 2017). In some instances, RNA is thought to seed NB assembly, and several NBs form at the sites of transcription, such as the nucleolus and paraspeckles. (Matera et al. 2009; Mao et al. 2011) A number of other seeding events exist (Dundr 2011; Mao et al. 2011; Shevtsov and Dundr 2011; White et al. 2011; Dellaire et al. 2006), and in many cases the precise mechanism of seeding formation is not known.

The HLB is present at RD histone genes even in G1 when the genes are not active, raising the possibility that histone genes themselves participate in seeding HLB assembly(Liu et al. 2006; Zhao et al. 1998). Consistent with this possibility, we previously identified a critical DNA sequence within the *Drosophila* histone locus that is necessary and sufficient for HLB assembly (Salzler et al. 2013). In *Drosophila*, the RD histone genes are present at a single locus with ~100 copies of a tandemly arrayed 5kb repeat unit, each containing one copy of the divergently transcribed *H2a-H2b* and *H3-H4* gene pairs as well as the gene for linker histone H1 (Lifton et al. 1978; McKay et al. 2015; Bongartz and Schloissnig 2018). Using transgenes containing a wild type or mutant derivatives of a single histone repeat, we demonstrated that sequences between the divergently transcribed *H3* and *H4* genes stimulated HLB assembly and transcription of the locus (Salzler et al. 2013). Thus, the *H3-H4* promoter region might act to seed HLB assembly. Within the *H3-H4* promoter region there are conserved GAGA repeats, and we demonstrated that a zinc-finger DNA binding protein, CLAMP, binds these repeats in vivo, and is present in the HLB(Rieder et al. 2017). In this work we use transgenic histone gene arrays to test whether the *H3-H4* promoter region is necessary for in vivo function of the RD histone

locus. We found that replacement of *H3-H4* promoters with *H2a-H2b* promoters results in an attenuated transgenic histone gene array that does not function in the presence of the intact endogenous histone locus but does provide full in vivo function when the endogenous histone locus is absent. These results suggest that histone genes compete for a limiting set of factors that nucleate HLB assembly.

MATERIALS AND METHODS

Culture condition and fly strains

Original fly strains and crosses were used as in McKay et al. 2015. Stocks were maintained on standard corn medium. Viability studies were performed as in Penke et al. 2016.

Locus Building

Construction of the 5kb histone repeat designed for this study was performed using NEB's HiFi DNA Assembly system. PCR amplification of fragments from existing histone repeats, in addition to IDT gblocks were employed for the building blocks of the reaction. Manufactures protocol was followed with slight modification to the incubation time of the reaction. The 5kb histone repeat designed was then arrayed to 12 copies in pMulti-BAC as in McKay et al. 2015 and integrated each into site VK33 on chromosome 3 using ϕC31-mediated integration (Model Systems Injections).

Northern Analysis

Northern analysis was performed using a 7 M 6% urea acrylamide gel to resolve histone mRNAs. ~1ug of RNA from embryos or larvae was used with a radio-labeled probe to the coding region of H3as in Lanzotti et al. 2002.

Histone Expression Analysis

Total RNA was prepared in Trizol and cDNA synthesized with random hexamers using Superscript II (Invitrogen), according to the manufacturer's instructions. RT-RCR was performed using gene-specific primers to H2a (McKay et al. 2015) and H3. PCR products were digested using XhoI (H2a) or SacI (H3). Digested amplicons were run on an 8% acrylamide gel or 1.5% agarose gel.

Immunofluorescence

We used primary antibodies at the following concentrations: rabbit anti-CLAMP (1:1000; giftfrom Leila Rieder), guinea pig anti-Mxc (1:2000), guinea pig anti-Mute (1:5000), rabbit anti-C terminus FLASH (1:2000), rabbit anti-Lsm10 (1:1000), mouse anti-MPM-2 (1:100; Millipore), rabbit anti-GAF (1:1000; gift from Leila Rieder), mouse anti-LacI (1:1000; Millipore). We used Alexa fluor secondary antibodies (Thermo Fisher Scientific) at a concentration of 1:1000. In situ probes were detected using 15 μg/mL streptavidin-DyLight-488 (Vector Laboratories). Salivary gland dissections and squashes were performed as in Tatomer et al. 2016. Images were acquired with z-stacks on Zeiss Lsm710 with ZEN DUO software. Images were analyzed using ImageJ. Ectopic HLBs in embryos were quantified as Salzler et al. 2016.

FISH-IF

FISH probe were made a PCR product that spanned the entire wild-type histone repeat in pUC57s as the template (primers F [AAAGGAGGTTGGTAGGCAGC] and R [ACGCTAGCGCTTTATCTGCA]).We made biotinylated FISH probes by nick translation using the purified PCR product: 1 μg of purified PCR product was incubated for 2 h at 15°C in a total of 50 μL containing 1×DNAPolI buffer (Fisher Optizyme); 0.05mMeach dCTP, dATP, and dGTP; 0.05 mM biotin-11-dUTP (Thermo Scientific); 10 mM 2-mercaptoethanol;0.004 U of

DNaseI (Fisher Optizyme); and 10 U of DNAPol I (Fisher Optizyme). The reaction was purified on a PCR purification column (Thermo Scientific) and diluted in hybridization buffer (2× SSC,10%dextran sulfate, 50%formamide, 0.8 mg/mL salmon sperm DNA) to a final volume of 220 μL. FISH probes were diluted in hybridization mixture and added to the slide before heating. We added a coverslip, sealed it with rubber cement, and heated the slide for 2 min on a 91°C heat block. Slides were placed in a humid box and incubated at 37C overnight. Immunostaining was then performed by incubating the slides in primary antibody overnight at 4C in a humid box.

RESULTS

To directly study histone gene regulation and the role of histone proteins we previously used BAC-based, transgenic histone gene arrays containing 12 copies of the histone 5kb repeat that assemble HLBs and functionally complement loss of the ~200 copy endogenous histone gene locus(McKay et al. 2015). We created a $12x^{HWT}$ (Histone Wild Type) transgenic construct containing a polymorphism in the *H2a* gene (i.e. mutation of a Xho I site) that allows us to distinguish transgenic histone gene expression from endogenous histone gene expression (McKay et al. 2015) . Here, we extended this design and created a wild type, $12x^{RDL}$ (Redesigned Locus) transgene that had all five histone genes marked in a similar manner (Fig. 3.2A). We also introduced restriction enzyme sites around the processing signals (the SL and HDE) to allow us to readily change those sequences (Fig. 3.2A). We can readily introduce changes into this designer repeat. We arrayed this RDL repeat into a $12x^{RDL}$ array. The $12x^{RDL}$ array rescued the lethality caused by deletion of the RD histone locus, resulting in viable, fertile adult flies that could be maintained as a stock.

The H3-H4 promoter stimulates HLB formation

To test whether the *H3-H4* promoter region is necessary for histone locus function in vivo, we engineered a derivative of the RDL repeat unit in which all *H3-H4* promoters were replaced with *H2a-H2b* promoters. In this 12x "Promoter Replacement" ($12x^{PR}$) construct we replaced the entire 298nt sequence between the initiation codons of the divergently transcribed *H3* and *H4* genes with the 226nt sequence between the initiation codons of the divergently transcribed *H2a* and *H2b* genes while leaving the *H1* and *H2a* and *H2b* genes intact (Fig. 1A). The *H2a-H2b* promoter region lacks the CLAMP-binding GAGA repeat elements of the *H3-H4* promoter and thus we reasoned that the $12x^{PR}$ construct would lack HLB nucleating sequences while retaining the ability to initiate transcription from all RD histone genes. The 72nt size difference between the promoters provides a way to unambiguously distinguish between the PR and RDL array genotypes using PCR (Fig. 3.2C).

We first assessed whether the $12x^{PR}$ and $12x^{RDL}$ transgene could support HLB formation in the presence of the endogenous histone genes in polytene chromosome spreads from 3rd instar larval salivary glands. In these polyploid cells the genome is amplified greater than 1000-fold and sister chromatids line up in register resulting in large chromosomes providing high-resolution cytology. We visualized the HLB by immunofluorescence using antibodies that recognize one of several HLB components. These components include Multi sex combs (Mxc), the *Drosophila* ortholog of human NPAT necessary for HLB assembly and histone gene expression (White et al. 2011; Terzo et al. 2015), histone pre-mRNA processing factors FLICE-associated huge protein (FLASH) and Lsm11, a component of the U7 snRNP (Burch et al. 2011; Yang et al. 2009), and Mute, a putative repressor of histone gene expression (Bulchand et al. 2010). In these preparations we also visualized both the endogenous histone locus and the ectopic, transgenic histone genes by FISH using a probe derived from the entire histone repeat

unit. Using these reagents, we observed HLB formation at the control $12x^{RDL}$ transgenic locus but not at the $12x^{PR}$ transgenic locus (Fig. 3.2D). These data indicate that sequences within the *H3-H4* promoter are necessary for HLB assembly in the presence of the endogenous histone genes.

Both single copy transgenic histone gene repeats that fail to form an HLB (Salzler et al. 2013) and Mxc mutants that don't form an HLB (Terzo et al. 2015) result in reduced histone mRNA levels, suggesting that HLB formation is necessary for efficient histone mRNA biosynthesis. We determined whether the $12x^{PR}$ transgene could support histone gene expression in the absence of HLB assembly. To determine expression from the ectopic $12x^{RDL}$ or $12x^{PR}$ genes in the presence of the endogenous RD histone genes, we randomly prime cDNA from total RNA and then amplify a fragment of each coding region containing the restriction enzyme site change. By digesting the PCR fragment with the appropriate restriction enzyme, and resolving the fragments by gel electrophoresis, we can determine the relative level of expression of each gene. For example, RT-PCR of the *H2a* gene results in the same size amplification product from both the transgenic and endogenous genes, but only the product from the endogenous histone mRNAs is sensitive to digestion with XhoI (Fig. 3.2B). When assayed in the presence of the endogenous genes, we found that *H2a* was expressed at high levels from $12x^{RDL}$ but not from $12x^{PR}$ transgene, even though identical promoters drove *H2a* in each transgenic array (Fig. 3.2B). Similar results were found for all five histone genes. This result is consistent with our previous observation using transgenes with a single histone gene repeat unit (Terzo et al. 2015; Salzler et al. 2013). Together these results demonstrate that the *H3-H4* promoter is required for both HLB formation and histone gene expression in the presence of the endogenous RD histone genes.

A histone gene array lacking the H3-H4 promoter forms HLBs and is expressed in the absence of the endogenous genes.

To determine the biological activity of the $12x^{PR}$ transgene directly, we determined the developmental outcome of having a $12x^{PR}$ transgene as the only zygotic source of histone mRNA. Due to large stores of maternal histone protein and mRNAs, embryos homozygous for a deletion that removes the endogenous histone gene array develop normally through S phase of cell cycle 14, but require zygotic RD histone gene expression for normal replication beginning in S phase of cycle 15 (Gunesdogan, Jackle, and Herzig 2010). Consequently, embryos lacking histone genes cannot complete cycle 15 and do not hatch. This embryonic lethality is rescued by a single $12x^{RDL}$ transgene, which supports development of histone deletion progeny into viable, fertile, adult flies. Surprisingly, we found that embryos lacking endogenous histone genes and containing the $12x^{PR}$ transgene hatched and developed into nearly the expected number of fertile adults without any overt developmental delays. These data suggest that the $12x^{PR}$ transgene provides normal amounts of histone gene function in the absence of the endogenous genes.

We interrogated this unexpected result more thoroughly by taking three complementary approaches (1) analyzing histone gene expression (2) HLB formation and (3) histone pre-mRNA processing in both 5-7 hour old embryos and 3rd instar larvae lacking endogenous histone genes and containing either $12x^{RDL}$ or $12x^{PR}$ transgenes. In both genotypes, all histone mRNA in either 5-7 hr embryos (Fig. 3.2G, lanes 3 and 5) or 3rd instar larvae (Fig. 3.2F, lanes 4 and 5) was derived only from the ectopic histone gene array. We observed robust histone gene expression from the $12x^{PR}$ transgene when the endogenous histone genes are absent, in stark contrast with the low level of expression in embryos from the $12x^{PR}$ transgene when endogenous histone genes are present (Fig. 3.2G, lane 4; Fig. 1F lane 3).

High levels of histone gene expression are strongly correlated with the ability to form an HLB (Salzler et al. 2013; Terzo et al. 2015; Rieder et al. 2017). Given that the $12x^{PR}$ can support histone gene expression in the absence of the endogenous genes, we assayed for HLB formation in polytene chromosome spreads from 3rd instar larval salivary glands and in embryos. We detected robust HLB formation at the $12x^{PR}$ transgenic locus with antibodies against FLASH, Lsm11 (Fig. 3.2H), Mxc, and Mute (not shown) similar to that observed at the RDL locus. Thus, the $12x^{PR}$ transgene, which lacks *H3-H4* promoter sequences, can support HLB formation and histone gene expression in the absence of endogenous histone genes.

Figure 3.2. The H3-H4 bidirectional promoter promotes but is not required for HLB formation.

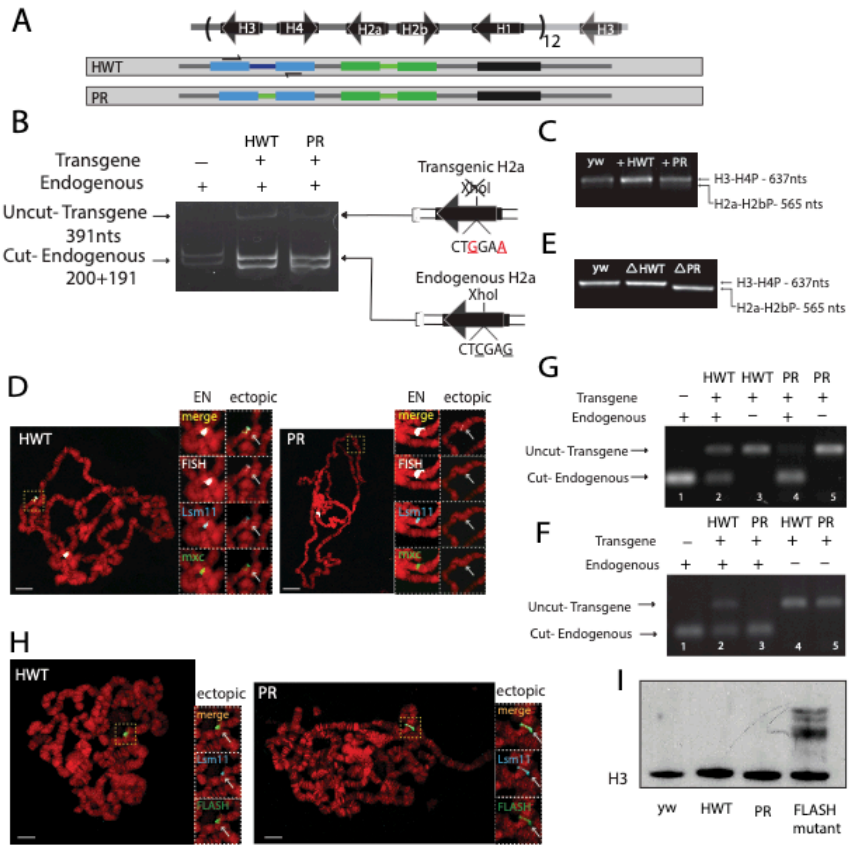


Figure 3.2. The H3-H4 bidirectional promoter promotes but is not required for HLB formation.

A.) Schematic of the BAC-based WT and PR synthetic histone repeats. Repeats were arrayed to 12 copies and inserted on chromosome 3. B.) Total RNA was prepared from 3rd instar larvae containing the endogenous genes, and one copy of either the WT or PR array. Silent mutations to the XhoI restriction site in the coding region of H2a were introduced into the synthetic histone arrays, destroying the restriction sites. RT-PCR analysis with H2a primers was performed on cDNA followed by digestion with XhoI, which will digest the cDNA from the endogenous genes but not the ectopic arrays and visualized with ethidium bromide on an 8% polyacrylamide gel. C. and E.) Genomic DNA was extracted from whole flies heterozygous (C) or homozygous (E) for a histone deletion. Primers that anneal to the coding regions of H3 and H4 (Arrow shown in A.) were used to amplify across the promoter region. The amplicon from a wild type locus is distinctly larger (637nts) than the PR locus (565nts). D) Polytene chromosome squashes from 3rd instar larval salivary glands from larvae that carried both the endogenous locus and histone transgenes were hybridized with a probe to the histone locus and stained for HLB components, Lsm11 and Mxc. HLBs are observed on the WT array, but in the presence of the endogenous genes HLB formation was never observed on the PR array. (G and F.) RT-PCR analysis with H2a primers performed on cDNA from 5-7 hr embryos (G) and 3rd instar larvae, (F) followed by digestion with XhoI, and visualized with ethidium bromide on a 1.5% agarose gel. H) Polytene chromosome spreads from 3rd instar larvae rescued by either the HWT or PR transgene were stained with antibodies against HLB components, Lsm11 and FLASH (I) Total RNA from 3rd instar larvae was analyzed by Northern blotting using a radiolabeled H3 coding region probe. A

FLASH mutant that cannot recruit the processing machinery was used as a positive control for production of polyA+ histone mRNA.

The *H3-H4* promoter is highly conserved among *drosophilids* and contains conserved GA repeats (Salzler et al. 2013; Rieder et al. 2017). We showed previously that these GAGA sequences are essential for HLB formation and expression of RD histone genes in the presence of the endogenous histone locus. Although in vitro these repeats can bind both the zinc-finger GA binding protein CLAMP (***c**hromatin **l**inked **a**daptor for **M**SL [**M**ale **S**pecific **L**ethal] **p**roteins*), and the major *Drosophila* GA-repeat binding protein GAF (*GAGA factor; trithorax-like (trl)*), only CLAMP is bound to the histone locus in wild-type animals.(Rieder et al. 2017). The H2a-H2b promoter (Fig. 3.3D) and the rest of the histone repeat unit do not contain any GA repeats longer than 4nts. We asked whether CLAMP is recruited to the PR transgene in salivary gland polytene chromosomes. As with all other HLB components we tested, CLAMP is not recruited to the PR transgene or a 12x histone gene array in which the GAGA sequences are replaced with *lacO* binding sites (Rieder et al. 2017) when the endogenous histone genes are present. Surprisingly, we found that in the absence of endogenous genes CLAMP (Fig. 3.3B, right), but not GAF (Fig. 3.3C, right), is recruited to the $12x^{PR}$ transgenic locus, with similar intensity to CLAMP recruitment to the $12x^{RDL}$ transgenic locus (Fig. 3.3 B,C, left). Furthermore, in this genotype the GAGA mutant array supports high level histone H2a gene expression (Fig. 3.3E). These data indicate that CLAMP can be recruited to a histone gene array lacking GAGA repeats when the preferred GAGA binding sites within the H3-H4 promoter at the endogenous locus are absent.

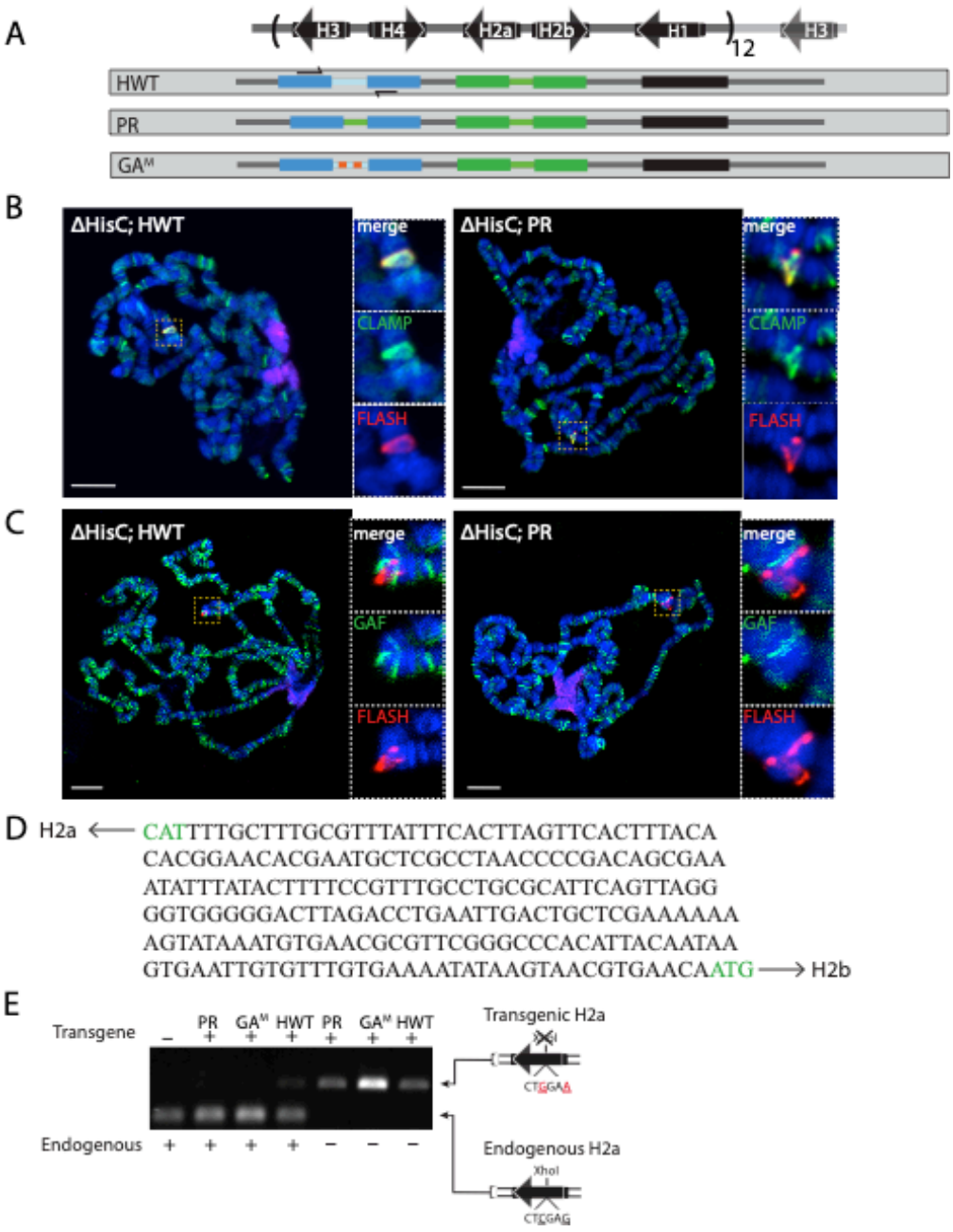


Figure 3.3. GA repeats are not required for CLAMP recruitment to the HLB

A. Schematic of the BAC-based WT, PR, and GA mutant (GA^M) synthetic histone repeats. For the GA^M , GA sequences in the H3-H4p were mutated to LacO sites or scrambled. Repeats were arrayed to 12 copies and inserted on chromosome 3. B. and C.) Polytene chromosome squashes from 3rd instar larvae rescued by a either a $12x^{RDL}$ or a $12x^{PR}$ transgene were stained with

antibodies against CLAMP (B) or GAF (C) and FLASH. D.) H2a-H2b bidirectional promoter sequence devoid of GA repeats. E.) RT-PCR analysis with H2a primers performed on cDNA from 5-7 hr embryos of indicated genotypes and visualized with ethidium bromide on a 0.8% agarose gel.

We carried out ChIP-qPCR experiments on embryos containing only the $12x^{RDL}$ array or the $12x^{PR}$ array, to determine whether CLAMP is interacting with either the *H2a-H2b* or *H3-H4* genes. In agreement with ChIP-seq results on the endogenous genes(Soruco et al. 2013), which showed CLAMP is localized precisely to the *H3-H4* promoter, CLAMP was bound to the H3-H4 promoter but not to the *H2a-H2b* gene or promoter on the WT array, and is not bound to either gene pair on the PR array, despite our observation that CLAMP is present in the HLB at the PR array. These data suggest that CLAMP can be recruited to histone genes by protein-protein interactions independently of its binding to DNA.

HLB assembly at the $12x^{PR}$ transgenic locus occurs at the normal time during embryogenesis

The above data suggests that the $12x^{PR}$ represents an attenuated histone gene array that provides normal biological function when not in competition with the wild type endogenous histone genes. To further explore this model, we determined if an HLB assembles on the $12x^{PR}$ array in the early embryo at the same time it assembles on the $12x^{RDL}$ array. The HLB begins assembling in syncytial blastoderm embryos just prior to the onset of zygotic histone transcription (White et al. 2011; Terzo et al. 2015). We previously reported that a "proto-HLB", consisting of Mxc and FLASH, forms in cycle 10 followed by the onset of zygotic histone gene expression and further recruitment of additional HLB components (Mute and Lsm11) in cycle 11 (Edgar and Schubiger 1986; White et al. 2011). To determine if the HLB forms at the $12x^{PR}$

transgenic locus with normal timing in early development, we stained syncytial blastoderm embryos lacking endogenous histone genes and rescued by a $12x^{PR}$ transgene with antibodies against Mxc (Fig. 3.4). In these experiments, HLB assembly during the syncytial blastoderm cycles were indistinguishable from that of histone deletion embryos rescued by the control $12x^{RDL}$ transgene. Thus, in the absence of the endogenous genes, an HLB assembles on the $12x^{PR}$ array at the same time as on the wild type array.

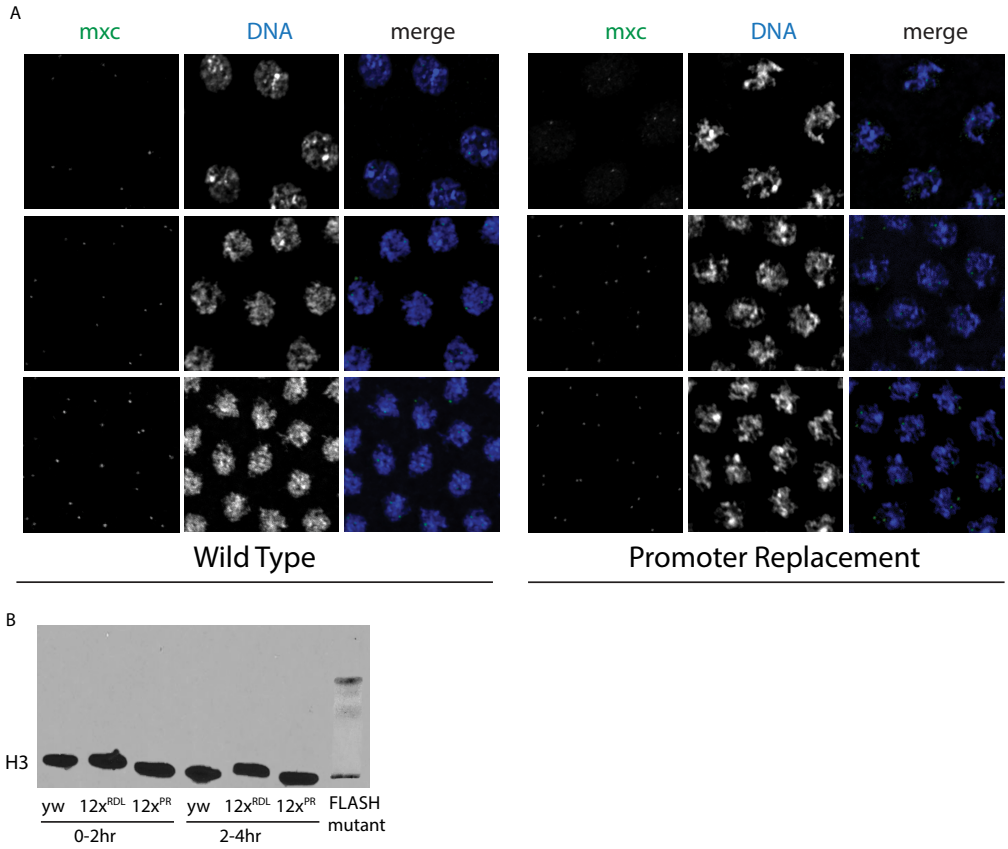


Figure 3.4 HLB formation is not delayed in PR embryos.

A. Early embryo collections from embryos recused by $12x^{RDL}$ or $12x^{PR}$ stained with Mxc to monitor HLB formation in development. B.) Total RNA was extracted from early embryos (0-2hrs) and 2-4hr collect, after zygotic genome activation takes place, and analyzed for proper histone processing using a Northern blot with a probe complementary to the H3 coding region.

Histone mRNA from the 12x^{PR} transgenic locus is properly processed

An important function of the HLB is concentrating factors to promote efficient histone pre-mRNA processing (Tatomer et al. 2016a). Therefore, we reasoned that it was possible that the attenuated PR gene array may affect other aspects of histone mRNA biosynthesis, including pre-mRNA processing.-All *Drosophila* RD histone genes contain cryptic polyadenylation signals downstream of the normal histone pre-mRNA processing sites that are only used when the histone processing reaction is compromised, resulting in the production and accumulation of poly(A)+ histone mRNA (Lanzotti et al. 2002; Godfrey et al. 2009; Tatomer et al. 2016b). Although we observed multiple HLB components localized to the 12x^{PR} transgenic locus, some may not be present in the correct stoichiometric amounts, resulting in misprocessed, poly(A)+ histone mRNA. We examined pre-mRNA processing efficiency in flies rescued by the 12x^{PR} transgene. Our PCR based assays to detect histone mRNA expression cannot differentiate between properly processed and misprocessed histone mRNAs. Therefore, we used northern blotting with an *H3* coding region probe to determine whether histone pre-mRNA was efficiently processed. In contrast to a *FLASH* mutant, which expresses large amounts of polyA+ histone mRNA, we did not detect poly(A)+ histone mRNA from early embryos (3.4 B) or from whole 3rd instar larvae (Fig. 3.2I) in histone locus deletion animals rescued by the 12x^{PR} transgene. These results indicate that the HLB formed on the 12x^{PR} array in the absence of endogenous histone genes supports efficient histone pre-mRNA processing.

A WT-array can activate the 12x^{PR} array when present in trans at the homologous locus

The results above demonstrate that the 12x^{PR} transgenic locus is a poor competitor for HLB components in the presence of the endogenous histone genes compared to the wild type

$12x^{RDL}$ transgene. We tested whether juxtaposing a wild type histone gene array near the $12x^{PR}$ locus would nucleate a functional HLB and activate expression from the $12x^{PR}$ transgene. Our BAC based transgenes are inserted into the genome via site-specific recombination at the same chromosomal location. We created flies in which the $12x^{PR}$ was placed in trans to the $12x^{HWT}$ transgene used in our initial studies at position VK33 on chromosome 3 (Fig. 3.5D), with the endogenous genes still present. We examined HLB formation at this ectopic location in the presence of the endogenous histone genes on chromosome 2 by staining intact salivary glands with antibodies against Mxc and FLASH. In 100% of the nuclei of this genotype, we observed formation of a second small HLB in addition to the single large endogenous HLB. By contrast, none of the nuclei contained an ectopic HLB in the presence of endogenous genes when $12x^{PR}$ was the only transgene present (Fig. 3.5B), consistent with the results of staining spread salivary gland chromosomes.

We next measured histone gene expression from the $12x^{PR}$ transgene in these genotypes. The $12x^{HWT}$ array used in this experiment has only the histone H2a gene marked with a restriction site change in the coding region (McKay et al. 2015), while the $12x^{PR}$ array has all five histone genes marked with a restriction enzyme change. We were able to specifically detect expression from the $12x^{PR}$ transgene by assaying the histone H3 mRNA using the SacI restriction enzyme site missing from the $12x^{PR}$ *H3* gene but present in both the endogenous and $12x^{HWT}$ *H3* genes (Fig. 3.5D). Strikingly, the $12x^{PR}$ transgene supports histone gene expression in embryos the presence of the endogenous histone genes when present in trans with the $12x^{HWT}$ transgene (Fig. 3.5D, lane 2). Together these data demonstrate that the wild type histone sequences in the $12x^{HWT}$ were able to activate the $12x^{PR}$ transgenic locus in the presence of the endogenous histone genes, likely by nucleating ectopic HLB formation that encompasses the paired

homologous chromosomes. This result is similar to transfection in *Drosophila*, which describes the phenomenon of gene activity being influenced by an enhancer in trans through pairing of homologous chromosomes (Duncan 2002).

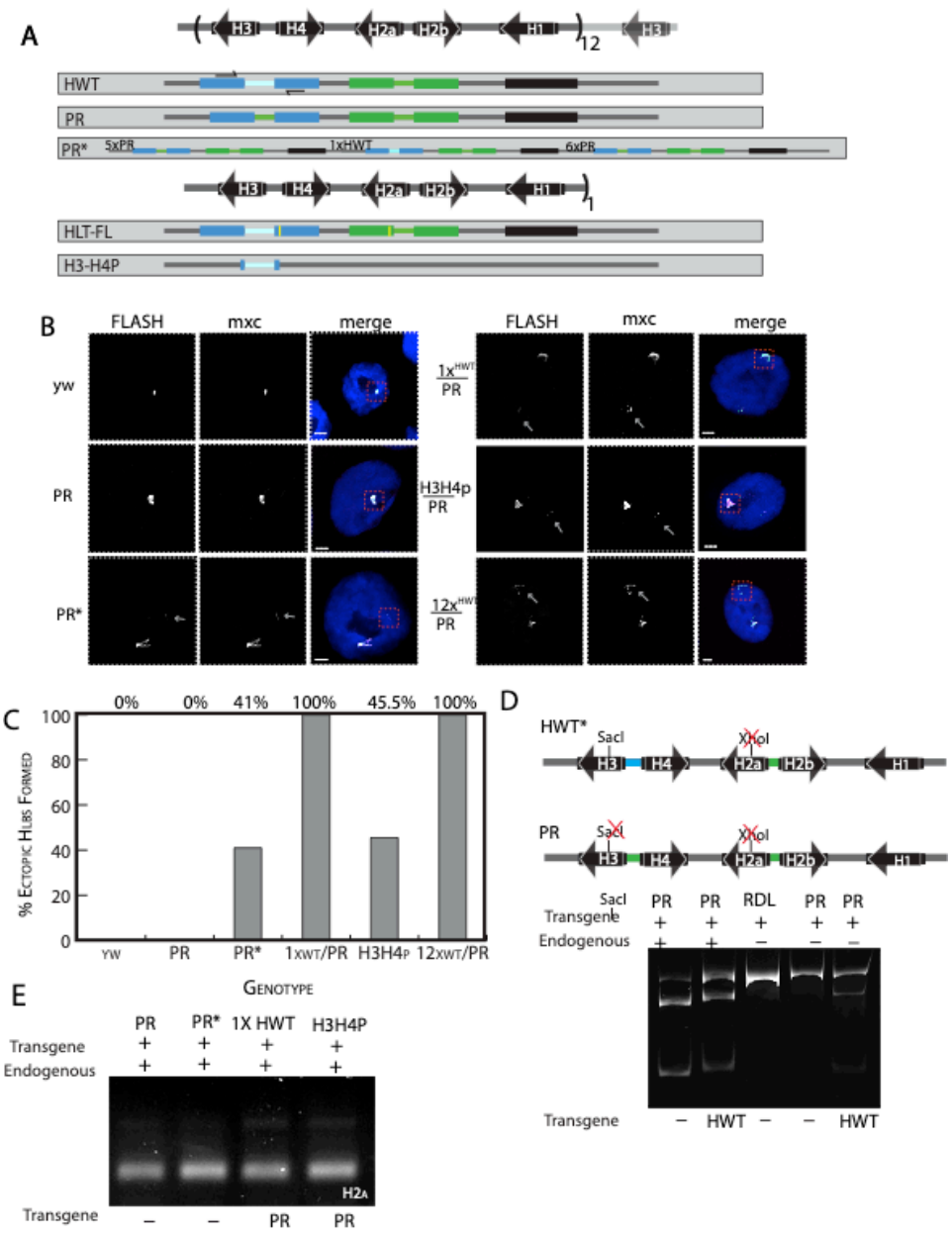


Figure 3.5 Activation of PR in trans

A. Schematic of the BAC-based WT, PR, and PR* synthetic histone repeats. Repeats were arrayed to 12 copies and inserted on chromosome 3. Below- Single copy full-length (HLT-FL)

histone repeat and *H3-H4p* histone transgenes. Transgenes were placed on chromosome 3. B.) Intact salivary gland nuclei stained for HLB components, Mxc and FLASH. C.) Quantification of the percentage of ectopic HLB formed by addition of wildtype histone gene repeats. D.) Schematic of the $12x^{PR}$ in trans to a $12x^{HWT}$. RT-PCR analysis with H3 primers performed on cDNA, followed by digestion with SacI from 5-7 hr embryos of indicated genotypes, and visualized with ethidium bromide on an 8% polyacrylamide gel. E.) RT-PCR analysis with H2a primers performed on cDNA, followed by digestion with XhoI from larvae of indicated genotypes, and visualized with ethidium bromide on an 8% polyacrylamide gel

Finding that 12 copies of a wild type histone repeat stimulated expression from the $12x^{PR}$ transgene, we tested whether a single copy could do the same. We observed that in the presence of the endogenous histone genes, a 1x HWT transgene induced HLB formation in 100% of the nuclei examined when placed in trans to $12x^{PR}$ (Fig. 3.5B). It also activated gene expression from $12x^{PR}$, albeit not as strongly as with $12x^{HWT}$ as qualitatively measured (Fig. 3.5E). We also observed HLB formation in 61% of nuclei when a transgene containing only one copy of the *H3-H4*p was placed in trans to the $12x^{PR}$ (Fig. 3.5B). Remarkably, the *H3-H4*p could also activate transcription from the $12x^{PR}$ transgene similar to the 1x HWT (Fig. 3.5E). Collectively, these data demonstrate that the presence of an HLB nucleating sequence in trans to $12x^{PR}$ can induce formation of an ectopic HLB and histone gene expression in the presence of the endogenous histone genes. Further, these data emphasize that the *H3-H4*p is a critical element in promoting HLB formation.

Considering that the $12x^{PR}$ can be activated in trans we were interested if a full-length wild type histone repeat in cis could stimulate the $12x^{PR}$ in a similar manner. We therefore

created a 12x array containing one wild-type full-length histone repeat in the center of 11 PR repeat units (the $12x^{PR*}$ construct) (Fig. 3.5A). Like $12x^{PR}$, the $12x^{PR*}$ transgene rescued a deletion of the endogenous histone genes, resulting in viable and fertile adults and indicating that it is likely fully active when present as the only source of RD histone genes. We then examined HLB formation in intact salivary glands from animals containing both the $12x^{PR*}$ transgene and the endogenous histone genes. In this genotype, we detected HLB formation at the ectopic $12x^{PR*}$ transgenic locus in 41% of the nuclei, compared to 0% of nuclei from the $12x^{PR}$ (Fig. 3.5D). A full-length histone repeat in cis could rescue HLB formation but we did not observe transcriptional activation in cis.

Discussion

Biomolecular condensates form when nucleating molecules seed and stabilize critical multi-component interactions that result in a high concentration of factors at a discrete location, resulting in a structure that can be observed by light microscopy (Gomes and Shorter 2018). Previously we provided evidence that sequences in the promoter region of the divergently transcribed *H3-H4* gene pair of the *Drosophila* histone locus is capable of nucleating HLB formation (Salzler et al. 2013). Here, we demonstrate that synthetic histone gene arrays completely lacking these sequences can form an HLB, but only if the endogenous genes have been deleted. There mutant histone genes arrays form an HLB at the normal time, express only properly processed histone mRNA and can fully support the entire *Drosophila* life cycle.

Nucleic acids, particularly RNA, have been proposed to nucleate many different NBs, although only the nucleolus and the HLB are associated with specific genomic loci (Mao et al. 2011; Shevtsov and Dundr 2011; Salzler et al. 2013). The activation of transcription of the

rRNA locus leads to the precise spatiotemporal formation of the nucleolus in *Drosophila* embryos (Falahati et al. 2016), and Men∈/β RNA, is a potent nucleator of paraspeckles (Mao et al. 2011; Shevtsov and Dundr 2011), which are not localized at the Men∈/β genes. In the absence of rDNA, *Drosophila* nucleolar components still form high concentration assemblies, but these are smaller, more numerous, and do not form at the same time in the early embryo as the nucleolus. *Drosophila* HLB components also form foci in the absence of the RD histone locus, suggesting that the HLB displays similar thermodynamically regulated phase behavior and that the histone genes are important for its assembly (White et al. 2007; Salzler et al. 2013).

In the present study we have presented evidence that histone arrays, devoid of either the *H3-H4* promoter, or missing the binding elements for CLAMP, can form a functional HLB, but only if the endogenous histone genes have been deleted. These HLBs (1) form at the normal time, (2) express only properly processed histone mRNAs, and (3) contain all the factors, including CLAMP, present in the wild-type HLBs. In contrast a functional HLB forms on a 12x WT array in the presence of the endogenous genes. This result also demonstrates that there is nothing adjacent to the histone gene locus which preferentially stimulates formation of an HLB on the endogenous genes.

We favor the possibility that the histone locus DNA nucleates the HLB in *Drosophila*, with the *H3-H4*p region being particularly important. We have shown that Mxc is likely the critical factor that together with histone genes seeds HLB formation, to activate histone gene expression. A truncation mutant of Mxc that fails to recruit histone pre-mRNA processing factors still forms an HLB and activates histone gene expression at sufficient levels to complete development. A possible interpretation of these results is that Mxc is recruited together with critical histone transcription factors at the histone genes to seed the HLB. Multiple interactions

between factors that bind the histone H3-H4 promoter (e.g. CLAMP) and the H2a-H2b promoter may provide the platform that recruits Mxc to initiate HLB formation.

Formation of an HLB on an exogenous array, requires that the array compete effectively with the endogenous array. If it does not, we hypothesize that the endogenous array sequesters a critical HLB component(s), likely including Mxc. We demonstrate here that *H3-H4*p can be replaced with *H2a-H2b*p, indicating that other sequences in the histone locus, perhaps *H2a-H2b*p itself, participate in nucleating HLB formation. However, these other sequences are not as potent as *H3-H4*p and are only used efficiently when *H3-H4*p is absent from the cell. These observations suggest a competition for limiting HLB components, such that when both the $12x^{PR}$ and endogenous histone genes are present, HLB nucleation preferentially occurs at the endogenous histone locus rather that at the $12x^{PR}$ transgenic locus, and critical available HLB factor(s) are sequestered.

Previous work suggests that not all scaffolds are equivalent in their ability to stimulate biomolecular condensates. In artificial systems, changes in scaffold stoichiometry, which can come from changes in valency, alter the recruitment of components (Banani et al. 2016). Further, mathematical modeling studies have revealed that scaffolds can nucleate distinct complexes when at different concentrations and this can qualitatively alter the output (Yang and Hlavacek 2011). Additionally, Rao and colleagues demonstrated that P-bodies can be formed in multiple ways through different protein-protein or protein-nucleic acid interactions, with different interactions predominating under different conditions (Rao and Parker 2017). Therefore, different nucleators of the HLB (i.e. the *H3-H4*p or other sequences in the locus) may not result in identical outcomes. Although we did not detect a genetic or molecular defect when the $12x^{PR}$ transgene rescued the loss of the endogenous histone genes, we speculate that if the HLB was

"stressed" in this situation, for example by mis-localization of a critical pre-mRNA processing factor as we have shown previously (Tatomer et al. 2016a), defects in histone mRNA biosynthesis may become apparent.

Interactions among multivalent proteins, or multivalent protein-nucleic acid interactions, are driving forces in the assembly of biomolecular condensates (Shin and Brangwynne 2017). Here we show that loss of the usual DNA binding site (i.e. the GAGA repeat in the H3-H4p) for the HLB component CLAMP is not essential for HLB formation in the absence of competing endogenous histone genes. Moreover, in this situation CLAMP is still recruited to the HLB that assembles at the $12x^{PR}$ transgenic locus. Although CLAMP may bind another sequence in the $12x^{PR}$ transgene, no other favorable GAGA repeats are identifiable. One possibility is that CLAMP is recruited by protein-protein interactions, perhaps via the multivalent protein Mxc/NPAT which is required for HLB assembly(White et al. 2011; Terzo et al. 2015). Thus, this evidence suggests that HLB formation is a summation of many interactions, and the breaking of any single one can be overcome by other multivalent interactions within the body. We propose that these observations support a role for multivalent interactions in HLB formation.

CHAPTER 4: DISCUSSION

The cell spatiotemporally regulates its biochemical reactions through the use of biomolecular condensates. It is assumed that these condensates, both in the nucleus and in the cytoplasm, have a function in the cell. However, if biomolecular condensates are present in cells to organize/facilitate biological functions that would not be, more or less, prompted efficiently, if at all, remains an open question. To understand a biological question of this magnitude, one must take a step back and rather than focusing solely on the output there needs to be an understanding of the beginning, how and what is required for these condensates form. Over the years we have learned about different mechanisms and requirements for formation. For example, we know that the paraspeckle forms on a long ncRNA, the nucleolus forms with the help of actively transcribed rRNA repeats, and the Cajal body forms with coilin as one of its structural components. However, the book is not closed on this question, rather these are just chapters. Another important question, one I eluded to earlier, is once formed do these structures provided an advantage to the cell. Do they establish an optimal environment for reactions, that otherwise wouldn't occur, to now happen efficiently? I have used the *Drosophila melanogaster* HLB to gain better insight into these questions.

The HLB is a powerful system to study this formation and function of condensates as we know where it forms and what processes are associated with it. In addition, working in *Drosophila* has enabled me to make tools to test my questions that I could not have otherwise made in other systems. In my thesis work I have approached the questions listed above by using genetic and biochemical approaches to understand what is required for the HLB to form

(Chapters 2 and 3) and further how the HLB functions in histone pre- mRNA processing (Appendix 1). Through my research I have shown one way in which the HLB is targeted to the histone locus and that the HLB can form via multiple different pathways. Further, I have shown that if an HLB cannot assemble on the locus, histone transcription is dramatically reduced, and this results in lethality. I have also demonstrated that by concentrating factors within the HLB histone pre-mRNA processing can occur efficiently (Appendix 1). Together these projects have highlighted how formation of a NB is linked to function

The GAGA sequences in the H3-H4 promoter triggers HLB assembly and biogenesis of histone mRNA.

The trans-acting factors necessary for HLB formation are understood relatively well (White et al. 2007; White et al. 2011; Wagner et al. 2007; Terzo et al. 2015). Through these studies, Mxc was identified as the scaffolding protein in the *Drosophila* HLB. Work from the field indicates that scaffolding proteins, with their many multivalent interactions, have the ability to regulate the composition and due to this, the function, of the body (Banani et al. 2016; Ditlev, Case, and Rosen 2018). Previous work has also told us that the HLB invariably forms at the replication-dependent histone locus (White et al. 2007; White et al. 2011; Rieder et al. 2017; Salzler et al. 2013). However, the scaffolding protein, Mxc, mentioned above, has not been shown to bind DNA (Terzo et al. 2015). Considering this, it is an open question as to how the HLB recognizes the histone locus. It seems likely that a sequence within the locus would bind a protein involved in histone metabolism to target components to the histone genes.

In Chapter 2 I had the pleasure to work with Drs. Leila Rieder and Erica Larschan to show that a protein involved in dosage compensation helped regulate the *Drosophila* histone

genes. This study defined important cis sequences as well as a trans factor governing HLB formation and function of the locus. First, we defined sequences within the *H3-H4* promoter that are essential for HLB formation. Second, we characterized a novel HLB factor, CLAMP, that bound to the GA sequences identified and functions to open the chromatin at the locus, enhance transcription, and promote HLB formation. By binding to these sequences, CLAMP can nucleate HLB components, as retargeting to the H3-H4 promoter in the absence of the GA sequences rescued HLB formation. Together these results suggest a way in which the HLB can recognize the histone locus and components can be targeted there. Before the onset of zygotic genome activation, a “proto-HLB” composed of FLASH and Mxc forms at the locus. However, as I referred to above, neither FLASH or Mxc has been shown to bind DNA. This leaves the possibility that a “pioneer” factor can bind the locus in early embryogenesis and initiates the formation of the "proto-HLB." It is possible that CLAMP is a member of the “proto-HLB.” These findings advance our knowledge of how the HLB is regulated, which is indispensable to our understanding of how the HLB functions in histone mRNA metabolism. However, questions about this property remain.

Things to test

When we retargeted CLAMP to the GA mutant array we created for this study, while HLB components were recruited, ectopic histone gene expression was not restored. Thus, HLB formation was uncoupled from histone transcription, something observed previously in the formation of the "proto-HLB." Furthermore, the morphology of the HLB was noticeably different. This could indicate that a threshold concentration of critical HLB components is needed to be passed for full HLB function. Several membraneless compartments exhibit a concentration threshold for assembly, and this is a hallmark for phase separation. (Shin and

Brangwynne 2017; Bolognesi et al. 2016). The HLB displays some phase separation properties, most notably, in the early embryo HLBs fuse, and the presence of intrinsically disordered regions (IDRs) in core proteins of the HLB, (Duronio and Marzluff 2017).

One possibility is that CLAMP-LacI cannot recruit critical components in high enough quantities or as efficiently as endogenous CLAMP. To test if CLAMP-LacI is unable to recruit components as efficiently as the endogenous CLAMP, a full-length CLAMP, $CLAMP^{FL}$-LacI, could be used to assay retargeted HLBs and if this is sufficient to activate transcription. If differences are observed when comparing the $CLAMP^{FL}$-LacI to the truncated CLAMP, $CLAMP^{\Delta zn}$-LacI, this may suggest that more factors are getting recruited or rather an essential factor is getting recruited by a full-length CLAMP. One would expect that if the full-length CLAMP construct could reach the endogenous CLAMP's ability to nucleate HLB components, then ectopic histone gene expression would be restored. In addition, the HLB could look morphologically different.

Knowing the number of components that are present in the HLB is an important question that has needed to be answered for quite some time, and this information is vital to our understanding of HLB function. There have been some very nice methods/protocols established to accurately and quantitively measure the protein composition of cellular complexes. A direct and straightforward way in which this can be done is by using ratios. Proteins can be counted by measuring the ratio of fluorescence intensity of a protein of interest, e.g., GFP-Mxc, to a standard with a known number of molecules. This method uses measurements from a series of images of cells that express either a fluorescently tagged protein of interest, or a tagged standard, but ideally, if the standard can be distinguished from the protein of interest, it is desirable to image cells that express both. (Coffman and Wu 2012). Using this method, protein components in the

HLB: Mxc, FLASH, Lsm10, Mute in addition to others could be counted at different stages which could reveal changes in HLB composition at different times (e.g., G1 vs. S phase). Also, it has been observed that HLBs formed on the endogenous histone repeat vs. our 12x transgenes are often different sizes. This same technique could be used to determine how the size of the proposed DNA "seed" or copy number, influences the number of components localized to in the HLB. Research by the Duronio lab and colleagues have generated a “gradient” of transgene sizes - 8x,12x,20x. These could be compared to a complete WT (100 or 200x) to determine if there exists a size scale mechanism.

Assembly of the HLB- looking for a “seed."

The assembly of the HLB has been best studied in Drosophila as the organization of the Drosophila histone genes are more amenable to manipulation and in addition, the organization of the human histone genes, with clusters on chromosome 1 and 6 with large distances in between, make it unlikely that there is a single sequence element that nucleates the HLB (Duronio and Marzluff 2017). Previous studies in the Duronio lab identified the *H3-H4* bidirectional promoter as the critical sequence directing HLB assembly and function. These studies were done in the presence of the endogenous histone genes as the single copy histone locus transgene the authors used was not intended to rescue viability (Salzler et al. 2013). Using engineered histone transgenes, I showed that only in the absence of the endogenous histone genes the H3-H4 promoter is not required for HLB assembly and function, but rather other sequences in the locus can nucleate a functional HLB. This result suggested that there was a competition for HLB components and that when the higher valency *H3-H4* promoter “seed” was present, it sequestered critical HLB from the PR. It also suggests that there is a limiting competent for HLB

assembly. One way in which this hypothesis can be tested is by overexpression of candidate HLB components and assaying for HLB formation and expression from the PR in the presence of endogenous genes. As Mxc is absolutely required HLB assembly (Terzo et al. 2015; White et al. 2011) it is reasonable that Mxc is this limiting component and overexpression of Mxc could rescue HLB formation and expression. In addition, it is conceivable that overexpressing CLAMP could recover HLB formation as CLAMP's preferred binding motifs are long GA repeats (Kuzu et al. 2016) of which are not found anywhere on the locus except the *H3-H4* promoter and CLAMP is only enriched on this promoter (Rieder et al. 2017). Considering this, a higher concentration of CLAMP could be needed to bind to the imperfect sites along the locus. This is analogous to RNA controlling phase transition behavior of RNA binding proteins, as has been seen for Whi3 (Zhang et al. 2015). Higher RNA concentration promotes aggregation and droplet formation at low concentrations of its RNA binding protein. Conversely, at low RNA concentrations, a much higher concentration of protein is needed for droplet formation (Zhang et al. 2015). This property comes from the increase in valency provided by the RNA. We favor a model in which the *H3-H4* promoter and the GA repeats provide a higher valency site to drive the formation of the HLB at lower protein concentration. To test this model, GAGA repeats could be put in the *H2a-H2b* promoter or elsewhere in the locus and assay for HLB formation and transcription in the presence of the endogenous histone repeat. If this does not stimulate HLB formation like the full-length *H3-H4* promoter, this would suggest other sequence elements are important for nucleating components. Mutation analysis, rather than deletions, which have been done in the past, of the *H3-H4* promoter, could be done to determine this.

I asked how many wild type promoters were required to make the PR competitive with the endogenous genes. In each case, I found that merely the presence of the *H3-H4* promoter in

trans to the PR stimulated HLB formation in salivary glands and moderately stimulated transcription. When we put the PR in trans to $12x^{HWT}$, more robust transcription was seen. This may suggest that increasing the number of repeats increases the number of HLB components recruited. We are assuming that the activation of the PR is from the HLB is being seeded off the *H3-H4* promoter and being "shared" between the histone loci as the transgene we used were at the same location and in Drosophila chromosomes pair (Duncan 2002).

To test if the activation of the PR in the presence of the endogenous genes was a result of chromosome pairing, the pairing of the transgenic histone loci could be prevented. We have done preliminary experiments to test this hypothesis. In an attempt to keep the PR transgene and the wild type transgene from pairing we used a transgene ($8x^{HWT}$) on the opposite chromosome arm then the PR (Appendix 2). Theoretically, when the chromosomes pair in this genotype, the transgenes would not be directly across from each other. HLB formation and histone gene expression were assayed in this genotype. We did not observe as much transcriptional activation as was seen with the $12x^{HWT}$ that was placed at the same site as the PR.

Further, when expression was assayed in the absence of endogenous genes, we observed robust expression from both the promoter replacement and the wildtype, suggesting that 8 copies of the *H3-H4* promoter are not enough to compete away available HLB components (Appendix 2). Future experiments will be aimed at increasing the number of histone repeats that cannot pair with the PR and determining if this can induce a shift in recruitment, presumably caused by the increased valency on the *H3-H4* promoter. This type of experiment has been done in artificial systems (Banani et al. 2016), and the ability to do this in vivo will be an excellent tool for the field. This would also highlight how multivalency serves to partition HLB components.

Robustness of the PR HLB

Because the PR HLB does not form on the usual seed (*H3-H4* promoter) is it is possible that it is not as functionally robust as a WT HLB. We have previously shown, by using FLASH mutants, that concentrating factors in the HLB is critical for proper processing. This resulted in uncoupling of 3' end processing and transcription termination (Tatomer et al. 2016a). In addition, a C-terminal truncation in mxc, mxc^{G46}, FLASH is mislocalized but is present at endogenous levels. This results in small amounts of misprocessed and read-through transcripts and this fly line is viable but female sterile. This mutant creates a situation in which the HLB is "stressed" measured by the mislocalization of one of the core components and the resulting alteration of the activity of the HLB (i.e., misprocessed histone transcripts). Putting the PR histone locus in a Mxc^{G46} background provides an ideal system for "stressing" the HLB. This provides a situation in which the HLB is "stressed" but viability is not affected at the starting point which would allow assessment of worsening phenotypes, if any, and effects on viability, among other things, could be efficiently measured when "stressing" the PR HLB. These experiments would add to the results we have so far on the PR and provide insight into whether the 'seed' influences the activity of the HLB.

The HLB is built on DNA

Work from the Dundr lab in LacO/LacI tethering experiments has gained popularity in the thinking that formation and structure of the HLB are dependent on histone mRNA production. This system is mostly artificial as tethering a component to a particular locus introduces a locally high concentration of components and this will more than likely lead to the formation of a cellular body, but it is clear that this is not physiologically relevant. We have

evidence that specific histone mRNA sequences are not required for HLB formation, but transcription initiation is required for subsequent recruitment of HLB factors. I hypothesize that it is a combination of both RNA and DNA that contribute to the structure we see in the microscope, but DNA is the 'seed,' rather than RNA. I propose an experiment that may give insight into these things.

In the lab, we can build our own histone locus, which enables us to directly determine the consequences of the mutation made to the locus in the absence of endogenous histone genes (McKay et al. 2015). I propose the construction of 2 new histone loci to assess the contribution if any, of RNA to the initial formation of the HLB.

Locus 1- This locus is designed to test if only histone RNA was the 'seed' for HLB assembly and RNA was required for the HLB to nucleate components. To construct this, we would use our 5kb repeat, but instead of having histone promoters (i.e., *H2a-H2*b bidirectional promoter or *H3-H4* bidirectional promoter) we would replace them with efficient nonhistone promoters, like actin. It is essential that the promoter be able to drive expression of the genes and that the nonhistone promoter had no similarities with the histone promoters as not to recruit HLB components due to sequence similarity. The histone coding regions would remain the same and because we have marked each gene (Figure 3.1) we can assess the expression of the genes.

Locus 2- This locus is designed to test if the HLB is seeded by the histone locus DNA, specifically the promoters, and formation and function are entirely independent of histone RNA. To construct this, the coding regions of the histone genes would be replaced by nonhistone coding regions. An additional locus could be created, one leaving off the histone processing signals (possibly replacing them with polyAs) and one which includes them. It is possible that the results between the two constructs would be different. Regardless, similar results to the *H3-*

H4 promoter only experiment (Salzler et al. 2013) may be observed, but it is possible of new proteins are recruited to the HLB, if there is one, built on this locus.

Locus 3- This would be the wild type histone locus which we would use as a positive control for HLB formation and function.

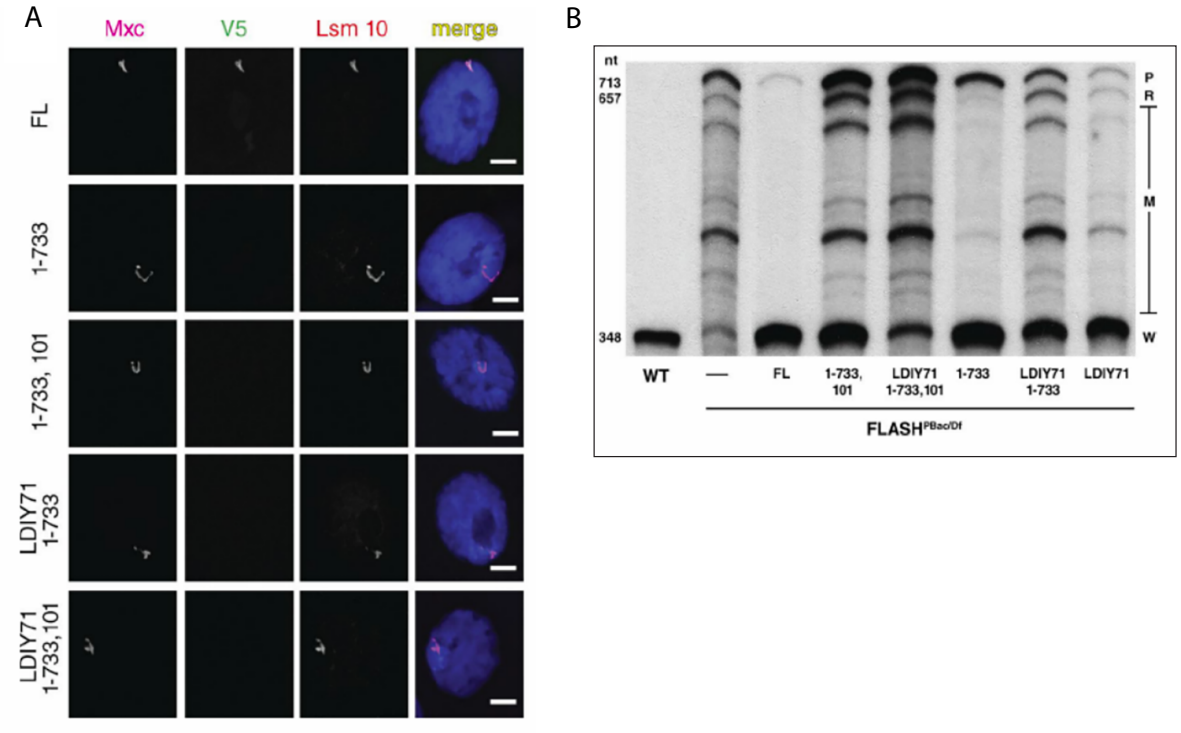
With all the constructed loci size, morphology, and timing of HLB construction in addition to histone mRNA biogenesis would be measured.

Complementary to this locus building would be purifying the HLB and treating it with RNase or DNase to completion, ensuring that all RNA or DNA is degraded. If the HLB is built on RNA, then destroying the RNA would have effects on the structure of the body and the same goes for DNA in DNase treating what is purified. If the methodology for purifying the HLB is robust, it would be interesting to determine the composition of the HLBs, if any, built on the experimental loci listed above.

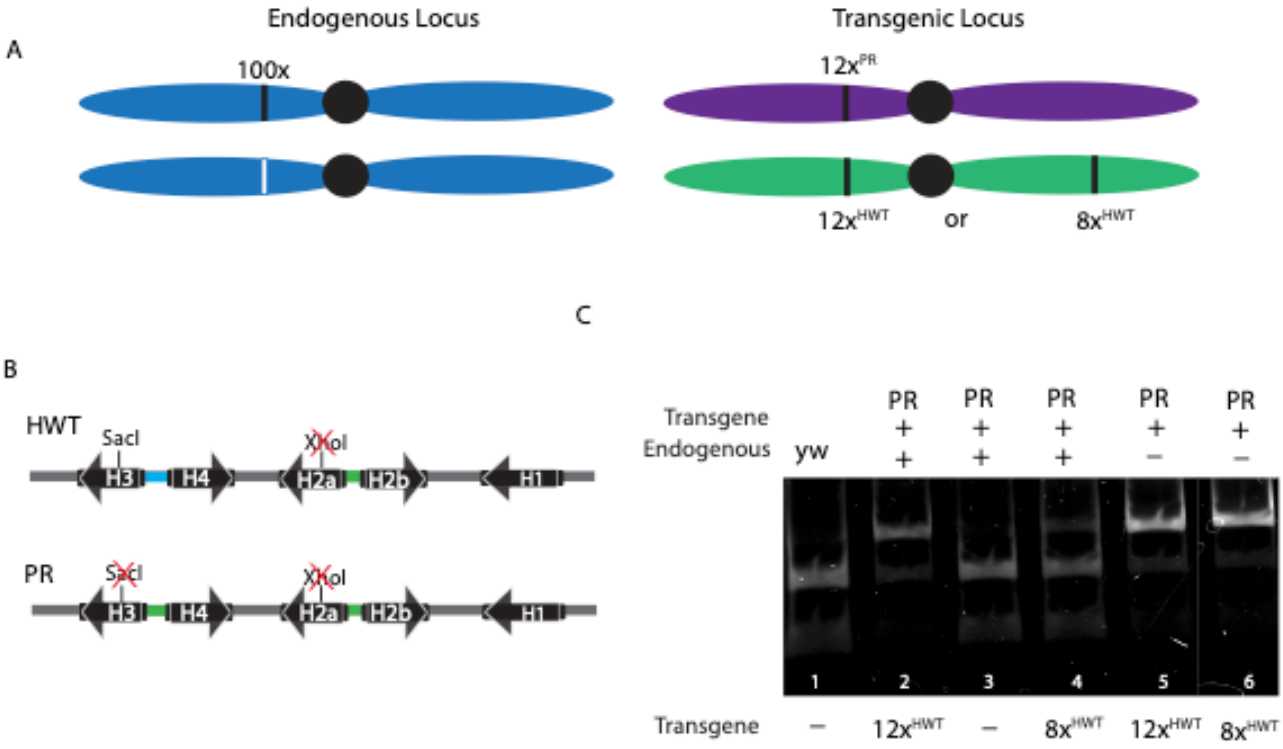
Many neurodegenerative diseases are marked by the accumulation of protein aggregates throughout the nervous system. RNA binding proteins (e.g., FUS, TDP-43, hnRNPA1) have been implicated in the pathological aggregates seen in ALS. Of interest, these proteins also play a role in the assembly stress granules (Elbaum-Garfinkle and Brangwynne 2015). This observation, and others, have stimulated extensive research efforts aimed at understanding the dysregulation of membraneless organelles and their contribution to disease. To appreciate the role of membraneless organelles in disease, there is a need for a thorough understanding of how these organelles form and function. The work presented here on the formation/function of the HLB is an important contribution to the efforts at understanding this fundamental question.

APPENDIX 2: ACTIAPPENDIX I: CONCENTRATING PRE-MRNA PROCESSING FACTORS IN THE HISTONE LOCUS BODY FACILITATES EFFICIENT HISTONE MRNA BIOGENESIS

Tatomer, D.C.,Terzo E, Curry K.P. et al. Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. J Cell Biol 213, 557-570 (2016



Appendix 1. Concentrating FLASH in the HLB promotes efficient histone pre-mRNA processing (A) Intact salivary glands from third-instar larvae expressing V5-tagged FLASH mutant proteins in FLASH PBac/Df were stained with anti-V5, anti-Lsm10, and anti-Mxc antibodies. Only the FL FLASH protein was detected in the HLBs in contrast to transgenes lacking the C-terminus indicating that the C-terminus facilitates HLB localization. Addition of the N-terminal 101 amino acids of Mxc, which is essential to target Mxc to the HLB, did not rescue HLB formation in this tissue. The C-terminus of FLASH is also required for U7 snRNP enrichment at the HLB. Lsm10 is observed in the HLBs in the FLASH PBac/Df mutants likely due to the perdurance of maternally loaded FLASH in salivary glands of third-instar larvae. (B). Total RNA from third-instar larvae expressing the transgenes in (A) was analyzed with a S1 nuclease protection assay using a 3'end labeled probe complementary to H2a mRNA that extends to the H4 HDE allowing detection of misprocessed H2a mRNA. The protected fragments were analyzed on a 6% polyacrylamide gel. Properly processed H2a mRNA protects a 348nt fragment from S1 nuclease digestion (W) and the cryptic PAS in H2a that longer, heterogenous populations of fragments (M). Together, these results indicate that concentrating pre-mRNA processing factors, FLASH and U7snRNP in the HLB, facilitate histone processing).



Appendix 2: To further characterize the ability of an HWT allele to activate the hypomorphic $12x^{PR}$ in the presence of the endogenous genes, a $12x^{HWT}$ or an $8x^{HWT}$ (A) was placed in trans to the PR transgene. We reasoned that if pairing of the chromosomes aided in activation of the $12x^{PR}$ by the $12x^{HWT}$, moving the HWT to a different arm on the chromosome would prevent pairing. (B) Using RT-PCR with gene specific primers to H3, followed by digestion with SacI, enabled differentiation of the HWT transgene and the PR transgene. (C) RNA from 3rd instar larvae of indicated genotypes was analyzed for H3 expression. In the presence of the endogenous genes (lanes 2, 3, 4), $12x^{HWT}$ at the same cis position as $12x^{PR}$ was able to activate histone gene expression from the $12x^{PR}$ to a higher level (lane 2) than an $8x^{HWT}$ located on the opposite arm of the chromosome (lane 4). Interestingly, in the absence of the endogenous genes, $12x^{PR}$ was expressed in the presence of $8x^{HWT}$ suggesting that 8-HWT copies was not enough to compete away HLB components from the PR.

REFERENCES

Albig, W., and D. Doenecke. 1997. 'The human histone gene cluster at the D6S105 locus', *Hum Genet*, 101: 284-94.

Altmeyer, M., K. J. Neelsen, F. Teloni, I. Pozdnyakova, S. Pellegrino, M. Grofte, M. B. Rask, W. Streicher, S. Jungmichel, M. L. Nielsen, and J. Lukas. 2015. 'Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose)', *Nat Commun*, 6: 8088.

Banani, S. F., H. O. Lee, A. A. Hyman, and M. K. Rosen. 2017. 'Biomolecular condensates: organizers of cellular biochemistry', *Nat Rev Mol Cell Biol*, 18: 285-98.

Banani, S. F., A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen. 2016. 'Compositional Control of Phase-Separated Cellular Bodies', *Cell*, 166: 651-63.

Belote, J. M., and J. C. Lucchesi. 1980. 'Control of X chromosome transcription by the maleless gene in Drosophila', *Nature*, 285: 573-5.

Berloco, M., L. Fanti, A. Breiling, V. Orlando, and S. Pimpinelli. 2001. 'The maternal effect gene, abnormal oocyte (abo), of Drosophila melanogaster encodes a specific negative regulator of histones', *Proc Natl Acad Sci U S A*, 98: 12126-31.

Bernues, J., D. Pineyro, and A. Kosoy. 2007. 'General, negative feedback mechanism for regulation of Trithorax-like gene expression in vivo: new roles for GAGA factor in flies', *Nucleic Acids Res*, 35: 7150-9.

Bhat, K. M., G. Farkas, F. Karch, H. Gyurkovics, J. Gausz, and P. Schedl. 1996. 'The GAGA factor is required in the early Drosophila embryo not only for transcriptional regulation but also for nuclear division', *Development*, 122: 1113-24.

Blythe, S. A., and E. F. Wieschaus. 2015. 'Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition', *Cell*, 160: 1169-81.

Bolognesi, B., N. Lorenzo Gotor, R. Dhar, D. Cirillo, M. Baldrighi, G. G. Tartaglia, and B. Lehner. 2016. 'A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression', *Cell Rep*, 16: 222-31.

Bond, C. S., and A. H. Fox. 2009. 'Paraspeckles: nuclear bodies built on long noncoding RNA', *J Cell Biol*, 186: 637-44.

Bongartz, P., and S. Schloissnig. 2018. 'Deep repeat resolution-the assembly of the Drosophila Histone Complex', *Nucleic Acids Res*.

Brangwynne, C. P., C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, and A. A. Hyman. 2009. 'Germline P granules are liquid droplets that localize by controlled dissolution/condensation', *Science*, 324: 1729-32.

Brangwynne, C. P., T. J. Mitchison, and A. A. Hyman. 2011. 'Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes', *Proc Natl Acad Sci U S A*, 108: 4334-9.

Bulchand, S., S. D. Menon, S. E. George, and W. Chia. 2010. 'Muscle wasted: a novel component of the Drosophila histone locus body required for muscle integrity', *J Cell Sci*, 123: 2697-707.

Burch, B. D., A. C. Godfrey, P. Y. Gasdaska, H. R. Salzler, R. J. Duronio, W. F. Marzluff, and Z. Dominski. 2011. 'Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila', *RNA*, 17: 1132-47.

Cakmakci, N. G., R. S. Lerner, E. J. Wagner, L. Zheng, and W. F. Marzluff. 2008. 'SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein', *Mol Cell Biol*, 28: 1182-94.

Chen, K., J. Johnston, W. Shao, S. Meier, C. Staber, and J. Zeitlinger. 2013. 'A global change in RNA polymerase II pausing during the Drosophila midblastula transition', *Elife*, 2: e00861.

Coffman, V. C., and J. Q. Wu. 2012. 'Counting protein molecules using quantitative fluorescence microscopy', *Trends Biochem Sci*, 37: 499-506.

Collier, S., A. Pendle, K. Boudonck, T. van Rij, L. Dolan, and P. Shaw. 2006. 'A distant coilin homologue is required for the formation of cajal bodies in Arabidopsis', *Mol Biol Cell*, 17: 2942-51.

Dellaire, G., C. H. Eskiw, H. Dehghani, R. W. Ching, and D. P. Bazett-Jones. 2006. 'Mitotic accumulations of PML protein contribute to the re-establishment of PML nuclear bodies in G1', *J Cell Sci*, 119: 1034-42.

Deryusheva, S., and J. G. Gall. 2009. 'Small Cajal body-specific RNAs of Drosophila function in the absence of Cajal bodies', *Mol Biol Cell*, 20: 5250-9.

Ditlev, J. A., L. B. Case, and M. K. Rosen. 2018. 'Who's In and Who's Out-Compositional Control of Biomolecular Condensates', *J Mol Biol*, 430: 4666-84.

Duncan, I. W. 2002. 'Transvection effects in Drosophila', *Annu Rev Genet*, 36: 521-56.

Dundr, M. 2011. 'Seed and grow: a two-step model for nuclear body biogenesis', *J Cell Biol*, 193: 605-6.

Dundr, M., and T. Misteli. 2010. 'Biogenesis of nuclear bodies', *Cold Spring Harb Perspect Biol*, 2: a000711.

Duronio, R. J., and W. F. Marzluff. 2017. 'Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body', *RNA Biol*, 14: 726-38.

Eagen, K. P., E. L. Aiden, and R. D. Kornberg. 2017. 'Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map', *Proc Natl Acad Sci U S A*, 114: 8764-69.

Edgar, Bruce A., and Gerold Schubiger. 1986. 'Parameters Controlling Transcriptional Activation during Early Drosophila Development', *Cell*, 44: 871-77.

Elbaum-Garfinkle, S., and C. P. Brangwynne. 2015. 'Liquids, Fibers, and Gels: The Many Phases of Neurodegeneration', *Dev Cell*, 35: 531-32.

Eriksson, P. R., D. Ganguli, V. Nagarajavel, and D. J. Clark. 2012. 'Regulation of histone gene expression in budding yeast', *Genetics*, 191: 7-20.

Falahati, H., B. Pelham-Webb, S. Blythe, and E. Wieschaus. 2016. 'Nucleation by rRNA Dictates the Precision of Nucleolus Assembly', *Curr Biol*, 26: 277-85.

Feric, M., N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne. 2016. 'Coexisting Liquid Phases Underlie Nucleolar Subcompartments', *Cell*, 165: 1686-97.

Fox, A. H., and A. I. Lamond. 2010. 'Paraspeckles', *Cold Spring Harb Perspect Biol*, 2: a000687.

Frey, M. R., and A. G. Matera. 1995. 'Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells', *Proc Natl Acad Sci U S A*, 92: 5915-19.

Fuda, N. J., M. J. Guertin, S. Sharma, C. G. Danko, A. L. Martins, A. Siepel, and J. T. Lis. 2015. 'GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters', *PLoS Genet*, 11: e1005108.

Gall, J. G. 2003. 'The centennial of the Cajal body', *Nat Rev Mol Cell Biol*, 4: 975-80.

Gemayel, R., S. Chavali, K. Pougach, M. Legendre, B. Zhu, S. Boeynaems, E. van der Zande, K. Gevaert, F. Rousseau, J. Schymkowitz, M. M. Babu, and K. J. Verstrepen. 2015. 'Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity', *Mol Cell*, 59: 615-27.

Gilmour, D. S., G. H. Thomas, and S. C. Elgin. 1989. 'Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters', *Science*, 245: 1487-90.

Godfrey, A. C., J. M. Kupsco, B. D. Burch, R. M. Zimmerman, Z. Dominski, W. F. Marzluff, and R. J. Duronio. 2006. 'U7 snRNA mutations in Drosophila block histone pre-mRNA processing and disrupt oogenesis', *RNA*, 12: 396-409.

Godfrey, A. C., A. E. White, D. C. Tatomer, W. F. Marzluff, and R. J. Duronio. 2009. 'The Drosophila U7 snRNP proteins Lsm10 and Lsm11 are required for histone pre-mRNA processing and play an essential role in development', *RNA*, 15: 1661-72.

Gomes, E., and J. Shorter. 2018. 'The molecular language of membraneless organelles', *J Biol Chem*.

Greenberg, A. J., J. L. Yanowitz, and P. Schedl. 2004. 'The Drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12DE', *Genetics*, 166: 279-89.

Groth, A. C., M. Fish, R. Nusse, and M. P. Calos. 2004. 'Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31', *Genetics*, 166: 1775-82.

Guertin, M. J., A. L. Martins, A. Siepel, and J. T. Lis. 2012. 'Accurate prediction of inducible transcription factor binding intensities in vivo', *PLoS Genet*, 8: e1002610.

Gunesdogan, U., H. Jackle, and A. Herzig. 2010. 'A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes', *EMBO Rep*, 11: 772-6.

Hamada, F. N., P. J. Park, P. R. Gordadze, and M. I. Kuroda. 2005. 'Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster', *Genes Dev*, 19: 2289-94.

Horard, B., C. Tatout, S. Poux, and V. Pirrotta. 2000. 'Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor', *Mol Cell Biol*, 20: 3187-97.

Hyman, A. A., and K. Simons. 2012. 'Cell biology. Beyond oil and water--phase transitions in cells', *Science*, 337: 1047-9.

Hyman, A. A., C. A. Weber, and F. Julicher. 2014. 'Liquid-liquid phase separation in biology', *Annu Rev Cell Dev Biol*, 30: 39-58.

Imai, Y., T. Kimura, A. Murakami, N. Yajima, K. Sakamaki, and S. Yonehara. 1999. 'The CED-4-homologous protein FLASH is involved in Fas-mediated activation of caspase-8 during apoptosis', *Nature*, 398: 777-85.

Kaiser, T. E., R. V. Intine, and M. Dundr. 2008. 'De novo formation of a subnuclear body', *Science*, 322: 1713-7.

Kasinathan, S., G. A. Orsi, G. E. Zentner, K. Ahmad, and S. Henikoff. 2014. 'High-resolution mapping of transcription factor binding sites on native chromatin', *Nat Methods*, 11: 203-9.

Kuzu, G., E. G. Kaye, J. Chery, T. Siggers, L. Yang, J. R. Dobson, S. Boor, J. Bliss, W. Liu, G. Jogl, R. Rohs, N. D. Singh, M. L. Bulyk, M. Y. Tolstorukov, and E. Larschan. 2016. 'Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites

on the X-Chromosome to Promote Drosophila Dosage Compensation', *PLoS Genet*, 12: e1006120.

Lanzotti, D. J., H. Kaygun, X. Yang, R. J. Duronio, and W. F. Marzluff. 2002. 'Developmental Control of Histone mRNA and dSLBP Synthesis during Drosophila Embryogenesis and the Role of dSLBP in Histone mRNA 3' End Processing In Vivo', *Molecular and Cellular Biology*, 22: 2267-82.

Larschan, E., M. M. Soruco, O. K. Lee, S. Peng, E. Bishop, J. Chery, K. Goebel, J. Feng, P. J. Park, and M. I. Kuroda. 2012. 'Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation', *PLoS Genet*, 8: e1002830.

Li, P., S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon, and M. K. Rosen. 2012. 'Phase transitions in the assembly of multivalent signalling proteins', *Nature*, 483: 336-40.

Lifton, R. P., M. L. Goldberg, R. W. Karp, and D. S. Hogness. 1978. 'The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications', *Cold Spring Harb Symp Quant Biol*, 42 Pt 2: 1047-51.

Liu, J. L., C. Murphy, M. Buszczak, S. Clatterbuck, R. Goodman, and J. G. Gall. 2006. 'The Drosophila melanogaster Cajal body', *J Cell Biol*, 172: 875-84.

Liu, Ji-Long, Zheng'an Wu, Zehra Nizami, Svetlana Deryusheva, T.K. Rajendra, Kelly J. Beumer, Hongjuan Gao, A. Gregory Matera, Dana Carroll, and Joseph G. Gall. 2009. 'Coilin Is Essential for Cajal Body Organization in Drosophila melanogaster', *Molecular and Cellular Biology*, 20.

Ma, T., B. A. Van Tine, Y. Wei, M. D. Garrett, D. Nelson, P. D. Adams, J. Wang, J. Qin, L. T. Chow, and J. W. Harper. 2000. 'Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription', *Genes Dev*, 14: 2298-313.

Machyna, M., K. M. Neugebauer, and D. Stanek. 2015. 'Coilin: The first 25 years', *RNA Biol*, 12: 590-6.

Mao, Y. S., H. Sunwoo, B. Zhang, and D. L. Spector. 2011. 'Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs', *Nat Cell Biol*, 13: 95-101.

Mao, Y. S., B. Zhang, and D. L. Spector. 2011. 'Biogenesis and function of nuclear bodies', *Trends Genet*, 27: 295-306.

Martin Muschol, and Franz Rosenberger. 1997. 'Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation:crystallization.', *American Institute of Physics*
, 107.

Marzluff, W. F., P. Gongidi, K. R. Woods, J. Jin, and L. J. Maltais. 2002. 'The human and mouse replication-dependent histone genes', *Genomics*, 80: 487-98.

Marzluff, W. F., and K. P. Koreski. 2017. 'Birth and Death of Histone mRNAs', *Trends Genet*, 33: 745-59.

Marzluff, W. F., E. J. Wagner, and R. J. Duronio. 2008. 'Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail', *Nat Rev Genet*, 9: 843-54.

Matera, A. G., M. Izaguire-Sierra, K. Praveen, and T. K. Rajendra. 2009. 'Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly?', *Dev Cell*, 17: 639-47.

McKay, D. J., S. Klusza, T. J. Penke, M. P. Meers, K. P. Curry, S. L. McDaniel, P. Y. Malek, S. W. Cooper, D. C. Tatomer, J. D. Lieb, B. D. Strahl, R. J. Duronio, and A. G. Matera. 2015. 'Interrogating the function of metazoan histones using engineered gene clusters', *Dev Cell*, 32: 373-86.

Mieczkowski, J., A. Cook, S. K. Bowman, B. Mueller, B. H. Alver, S. Kundu, A. M. Deaton, J. A. Urban, E. Larschan, P. J. Park, R. E. Kingston, and M. Y. Tolstorukov. 2016. 'MNase titration reveals differences between nucleosome occupancy and chromatin accessibility', *Nat Commun*, 7: 11485.

Miele, A., C. D. Braastad, W. F. Holmes, P. Mitra, R. Medina, R. Xie, S. K. Zaidi, X. Ye, Y. Wei, J. W. Harper, A. J. van Wijnen, J. L. Stein, and G. S. Stein. 2005. 'HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition', *Mol Cell Biol*, 25: 6140-53.

Misteli, T. 2001. 'The concept of self-organization in cellular architecture', *J Cell Biol*, 155: 181-5.

Mitrea, D. M., and R. W. Kriwacki. 2016. 'Phase separation in biology; functional organization of a higher order', *Cell Commun Signal*, 14: 1.

Mowry, K. L., and J. A. Steitz. 1987. 'Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's', *Science*, 238: 1682-7.

Ni, J. Q., M. Markstein, R. Binari, B. Pfeiffer, L. P. Liu, C. Villalta, M. Booker, L. Perkins, and N. Perrimon. 2008. 'Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster', *Nat Methods*, 5: 49-51.

Novotny, I., M. Blazikova, D. Stanek, P. Herman, and J. Malinsky. 2011. 'In vivo kinetics of U4/U6.U5 tri-snRNP formation in Cajal bodies', *Mol Biol Cell*, 22: 513-23.

O'Brien, T., R. C. Wilkins, C. Giardina, and J. T. Lis. 1995. 'Distribution of GAGA protein on Drosophila genes in vivo', *Genes Dev*, 9: 1098-110.

Ohtsuki, S., and M. Levine. 1998. 'GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo', *Genes Dev*, 12: 3325-30.

Pandey, N. B., and W. F. Marzluff. 1987. 'The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability', *Mol Cell Biol*, 7: 4557-9.

Paris, M., T. Kaplan, X. Y. Li, J. E. Villalta, S. E. Lott, and M. B. Eisen. 2013. 'Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression', *PLoS Genet*, 9: e1003748.

Patel, A., H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, and S. Alberti. 2015. 'A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation', *Cell*, 162: 1066-77.

Pederson, T. 2011. 'The nucleolus', *Cold Spring Harb Perspect Biol*, 3.

Quinn, J. J., I. A. Ilik, K. Qu, P. Georgiev, C. Chu, A. Akhtar, and H. Y. Chang. 2014. 'Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification', *Nat Biotechnol*, 32: 933-40.

Rajendra, T. K., K. Praveen, and A. G. Matera. 2010. 'Genetic analysis of nuclear bodies: from nondeterministic chaos to deterministic order', *Cold Spring Harb Symp Quant Biol*, 75: 365-74.

Rao, B. S., and R. Parker. 2017. 'Numerous interactions act redundantly to assemble a tunable size of P bodies in Saccharomyces cerevisiae', *Proc Natl Acad Sci U S A*, 114: E9569-E78.

Rieder, L. E., K. P. Koreski, K. A. Boltz, G. Kuzu, J. A. Urban, S. K. Bowman, A. Zeidman, W. T. Jordan, 3rd, M. Y. Tolstorukov, W. F. Marzluff, R. J. Duronio, and E. N. Larschan. 2017. 'Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP', *Genes Dev*, 31: 1494-508.

Robinett, C. C., A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, and A. S. Belmont. 1996. 'In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition', *J Cell Biol*, 135: 1685-700.

Russo, C. A., N. Takezaki, and M. Nei. 1995. 'Molecular phylogeny and divergence times of drosophilid species', *Mol Biol Evol*, 12: 391-404.

Saget, O., F. Forquignon, P. Santamaria, and N. B. Randsholt. 1998. 'Needs and targets for the multi sex combs gene product in Drosophila melanogaster', *Genetics*, 149: 1823-38.

Salzler, H. R., D. C. Tatomer, P. Y. Malek, S. L. McDaniel, A. N. Orlando, W. F. Marzluff, and R. J. Duronio. 2013. 'A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs', *Dev Cell*, 24: 623-34.

Sawyer, I. A., and M. Dundr. 2016. 'Nuclear bodies: Built to boost', *J Cell Biol*, 213: 509-11.

Schienman, J. E., E. R. Lozovskaya, and L. D. Strausbaugh. 1998. 'Drosophila virilis has atypical kinds and arrangements of histone repeats', *Chromosoma*, 107: 529-39.

Shevtsov, S. P., and M. Dundr. 2011. 'Nucleation of nuclear bodies by RNA', *Nat Cell Biol*, 13: 167-73.

Shin, Y., and C. P. Brangwynne. 2017. 'Liquid phase condensation in cell physiology and disease', *Science*, 357.

Smith, M. M., and K. Murray. 1983. 'Yeast H3 and H4 histone messenger RNAs are transcribed from two non-allelic gene sets', *J Mol Biol*, 169: 641-61.

Soruco, M. M., J. Chery, E. P. Bishop, T. Siggers, M. Y. Tolstorukov, A. R. Leydon, A. U. Sugden, K. Goebel, J. Feng, P. Xia, A. Vedenko, M. L. Bulyk, P. J. Park, and E. Larschan. 2013. 'The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation', *Genes Dev*, 27: 1551-6.

Staller, M. V., D. Yan, S. Randklev, M. D. Bragdon, Z. B. Wunderlich, R. Tao, L. A. Perkins, A. H. Depace, and N. Perrimon. 2013. 'Depleting gene activities in early Drosophila embryos with the "maternal-Gal4-shRNA" system', *Genetics*, 193: 51-61.

Staněk, David, and Archa H. Fox. 2017. 'Nuclear bodies: news insights into structure and function', *Current Opinion in Cell Biology*, 46: 94-101.

Stroberg, W., and S. Schnell. 2018. 'Do Cellular Condensates Accelerate Biochemical Reactions? Lessons from Microdroplet Chemistry', *Biophys J*, 115: 3-8.

Strub, K., and M. L. Birnstiel. 1986. 'Genetic complementation in the Xenopus oocyte: co-expression of sea urchin histone and U7 RNAs restores 3' processing of H3 pre-mRNA in the oocyte', *EMBO J*, 5: 1675-82.

Strzelecka, M., S. Trowitzsch, G. Weber, R. Luhrmann, A. C. Oates, and K. M. Neugebauer. 2010. 'Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis', *Nat Struct Mol Biol*, 17: 403-9.

Sullivan, E., C. Santiago, E. D. Parker, Z. Dominski, X. Yang, D. J. Lanzotti, T. C. Ingledue, W. F. Marzluff, and R. J. Duronio. 2001. 'Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression', *Genes Dev*, 15: 173-87.

Sullivan, K. D., T. E. Mullen, W. F. Marzluff, and E. J. Wagner. 2009. 'Knockdown of SLBP results in nuclear retention of histone mRNA', *RNA*, 15: 459-72.

Tariq, M., R. Wegrzyn, S. Anwar, B. Bukau, and R. Paro. 2013. 'Drosophila GAGA factor polyglutamine domains exhibit prion-like behavior', *BMC Genomics*, 14: 374.

Tatomer, D. C., E. Terzo, K. P. Curry, H. Salzler, I. Sabath, G. Zapotoczny, D. J. McKay, Z. Dominski, W. F. Marzluff, and R. J. Duronio. 2016a. 'Concentrating pre-mRNA

processing factors in the histone locus body facilitates efficient histone mRNA biogenesis', *J Cell Biol*, 213: 557-70.

———. 2016b. 'Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis', *J Cell Biol*.

Taylor, J. P., R. H. Brown, Jr., and D. W. Cleveland. 2016. 'Decoding ALS: from genes to mechanism', *Nature*, 539: 197-206.

Terzo, E. A., S. M. Lyons, J. S. Poulton, B. R. Temple, W. F. Marzluff, and R. J. Duronio. 2015. 'Distinct self-interaction domains promote Multi Sex Combs accumulation in and formation of the Drosophila histone locus body', *Mol Biol Cell*, 26: 1559-74.

Tsai, S. Y., Y. L. Chang, K. B. Swamy, R. L. Chiang, and D. H. Huang. 2016. 'GAGA factor, a positive regulator of global gene expression, modulates transcriptional pausing and organization of upstream nucleosomes', *Epigenetics Chromatin*, 9: 32.

Tsukiyama, T., and C. Wu. 1995. 'Purification and properties of an ATP-dependent nucleosome remodeling factor', *Cell*, 83: 1011-20.

Urban, J. A., C. A. Doherty, W. T. Jordan, 3rd, J. E. Bliss, J. Feng, M. M. Soruco, L. E. Rieder, M. A. Tsiarli, and E. N. Larschan. 2017. 'The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females', *Chromosome Res*, 25: 101-13.

Wagner, E. J., B. D. Burch, A. C. Godfrey, H. R. Salzler, R. J. Duronio, and W. F. Marzluff. 2007. 'A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing', *Mol Cell*, 28: 692-9.

Walker, M. P., L. Tian, and A. G. Matera. 2009. 'Reduced viability, fertility and fecundity in mice lacking the cajal body marker protein, coilin', *PLoS One*, 4: e6171.

Weber, J. A., and D. S. Gilmour. 1995. 'Genomic footprinting of the hsp70 and histone H3 promoters in Drosophila embryos reveals novel protein-DNA interactions', *Nucleic Acids Res*, 23: 3327-34.

Wei, Y., J. Jin, and J. W. Harper. 2003. 'The cyclin E/Cdk2 substrate and Cajal body component p220(NPAT) activates histone transcription through a novel LisH-like domain', *Mol Cell Biol*, 23: 3669-80.

White, A. E., B. D. Burch, X. C. Yang, P. Y. Gasdaska, Z. Dominski, W. F. Marzluff, and R. J. Duronio. 2011. 'Drosophila histone locus bodies form by hierarchical recruitment of components', *J Cell Biol*, 193: 677-94.

White, A. E., M. E. Leslie, B. R. Calvi, W. F. Marzluff, and R. J. Duronio. 2007. 'Developmental and cell cycle regulation of the Drosophila histone locus body', *Mol Biol Cell*, 18: 2491-502.

Woulfe, J. 2008. 'Nuclear bodies in neurodegenerative disease', *Biochim Biophys Acta*, 1783: 2195-206.

Yang, J., and W. S. Hlavacek. 2011. 'Scaffold-mediated nucleation of protein signaling complexes: elementary principles', *Math Biosci*, 232: 164-73.

Yang, X. C., B. D. Burch, Y. Yan, W. F. Marzluff, and Z. Dominski. 2009. 'FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs', *Mol Cell*, 36: 267-78.

Yang, X. C., I. Sabath, J. Debski, M. Kaus-Drobek, M. Dadlez, W. F. Marzluff, and Z. Dominski. 2013. 'A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3'-end processing', *Mol Cell Biol*, 33: 28-37.

Yang, X. C., I. Sabath, L. Kunduru, A. J. van Wijnen, W. F. Marzluff, and Z. Dominski. 2014. 'A conserved interaction that is essential for the biogenesis of histone locus bodies', *J Biol Chem*, 289: 33767-82.

Ye, X., Y. Wei, G. Nalepa, and J. W. Harper. 2003. 'The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells', *Mol Cell Biol*, 23: 8586-600.

Zhang, H., S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, and A. S. Gladfelter. 2015. 'RNA Controls PolyQ Protein Phase Transitions', *Mol Cell*, 60: 220-30.

Zhao, J., B. Dynlacht, T. Imai, T. Hori, and E. Harlow. 1998. 'Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S- phase entry', *Genes Dev*, 12: 456-61.

Zhao, J., B. K. Kennedy, B. D. Lawrence, D. A. Barbie, A. G. Matera, J. A. Fletcher, and E. Harlow. 2000. 'NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription', *Genes Dev*, 14: 2283-97.

Zhu, L., and C. P. Brangwynne. 2015. 'Nuclear bodies: the emerging biophysics of nucleoplasmic phases', *Curr Opin Cell Biol*, 34: 23-30.